APRIL 1955
Volume XI Number 2

# Journal of Clinical Psychology

EDITORIAL BOARD

FREDERICK C. THORNE, *Editor*
Editorial Office: 5 Pearl Street
Brandon, Vermont

JERRY W. CARTER, JR.
*U. S. Public Health Service*

HOWARD F. HUNT
*University of Chicago*

WILLIAM A. HUNT
*Northwestern University*

ELAINE F. KINDER
*New York State Department of Mental Hygiene*

C. M. LOUTTIT
*Wayne University*

ROBERT I. WATSON
*Northwestern University*

ELEANOR C. THORNE, *Business Manager*

Journal of Clinical Psychology
5 Pearl Street - - Brandon, Vermont

## Principles of Psychological Examining

by FREDERICK C. THORNE, M.D. Ph.D.
Editor, Journal of Clinical Psychology

- The First Comprehensive Textbook of Applied Integrative Psychology.
- Systematic: Presenting a Complete System for the Differential Diagnosis of the Primary Etiologic Factors Which Actually Organize Socially Significant Behaviors.
- Complete: More Extensive Diagnostic Classifications of Etiologic Factors.
- Eclectic: A theoretical system is outlined for relating all pertinent theories and facts in basic science psychology and clinical psychological science.
- Advanced: More than a textbook of psychometrics. Tests and measures are presented only as they relate to more fundamental problems of psychological diagnosis.

**READY ABOUT JUNE 1955 - - PRICE: $7.50**

**PRE-PUBLICATION OFFER TO APA MEMBERS $6.00**

**ORDER NOW**

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

## THE HOWARD INK BLOT TEST

A NEW PROJECTIVE TECHNIQUE

- Derived experimentally by James W. Howard, Ph.D.
- An individual test consisting of 12 plates 8½" x 11". The blots are somewhat larger than the Rorschach and with a wider range of stimulating features. Blots of varied colors selected from an experimental sample of 10,000 blots.
- Not intended to be a parallel test to the Rorschach. Statistical findings from a "normal" group show the test to differ markedly from the Rorschach in ways which should indicate greater sensitivity.
- Selection and arrangement of blots made on the basis of evidence of character, degree and spread of stimulation.
- Data from normative sample indicate more frequent light determined responses, higher incidence of color responses with different C, CF and FC relationships, more movement responses and fewer animal responses. Greater diagnostic sensitivity with wider range of response.

**Price: $12.50** (In U. S. and Canada; foreign postage $.50 extra)

## HOWARD INK BLOT TEST MANUAL

MONOGRAPH SUPPLEMENT NO. 10 FROM JULY 1953 ISSUE

**Price $2.00**

JOURNAL OF CLINICAL PSYCHOLOGY
5 Pearl Street Brandon, Vermont

# JOURNAL OF CLINICAL PSYCHOLOGY

VOL. XI APRIL 1955 No. 2

## CONTENTS

Page

The Krout Personal Preference Scale: A Factor Analytic Study. Ross Stagner, Edwin D. Lawson, and J. Weldon Moffitt, University of Illinois ..... 103

Psychology Internship Training: A Panel Discussion. University of Houston, Texas ..... 114

Rater Experience and Case History Judgments of Adjustment. A. W. Bendig, University of Pittsburgh ..... 127

An Empirical Investigation of Therapeutic "Listening". G. R. Sommer, B. Mazo and G. F. J. Lehner, University of California, Los Angeles ..... 132

Estimating Impairment on Wechsler Bellevue Subtests. C.V. Jackson, National Hospital for Nervous Diseases, London ..... 137

The Expressive Value of Toys Recommended for Nondirective Play Therapy. Dell Lebo, Carter Memorial Hospital, Indianapolis ..... 144

An Item Analysis of the Wechsler Intelligence Scale for Children. Frederick O. Carleton and Chalmers L. Stacey, Syracuse University ..... 149

Use of the WISC in Predicting Organicity. Harry S. Beck, Illinois Department of Public Instruction, and Robert L. Lam, Washington University Medical School ..... 154

Teacher's Judgments of the Popularity of Children from their Human Figure Drawings. Alexander Tolor, Columbia-Presbyterian Medical Center, New York City ..... 158

Consistency of Reproductions of Bender-Gestalt Designs. Marion W. McPherson, Neuropsychiatric Clinic of St. Louis, and Loretta A. Pepin, Detroit Public Schools ..... 163

Inter-Rater Agreement in Judging Student Adjustment from Projective Tests. Ija N. Korner and Dale Westwood, University of Utah College of Medicine ..... 167

The Effect of Various Interview Techniques in Evoking Fear Responses. Charles Windle, Howard McFann and Joseph Ward, Fort Benning, Georgia ..... 171

A Factor Analytic Study of the Adjustment of Chronic Schizophrenics. Wilson H. Guertin, Veterans Administration Hospital, Knoxville, Iowa ..... 174

Hypohidrosis: A Case Study. J. L. McCary, University of Houston and James A. Brown, Almeda Clinic, Houston, Texas ..... 177

(Over)

CONTENTS—Continued

Mental Patients' Attitudes toward Mental Hospitals. Omneya Soulem, Cairo, Egypt ..... 181

The Bender-Gestalt Recall as an Index of Intellectual Functioning. Roland M. Peek and Gordon W. Olson, Hastings (Minn.) State Hospital ..... 185

Empathy or Projection? Richard C. Cowden, Veterans Administration Hospital, Gulfport, Miss. ..... 188

Scoring Colored Responses in Paintings. John A. Popplestone, Washington University ..... 191

Rorschach and Levy Movement Responses: A Research Note. Gerald F. King, Veterans Administration Hospital, Battle Creek, Mich. ..... 193

An Aid in Calculating Q-Sort Factor-Arrays. James W. Creaser, University of Illinois ..... 195

Historical Inaccuracies. J. W. W. Wallin, Wilmington, Delaware ..... 197

Editorial Opinion ..... 201

Book Notices ..... 203

---

## GENERAL INFORMATION

The *Journal of Clinical Psychology* is an independent journal dedicated to the advancement of the clinical method in psychology. Although primarily a scientifically oriented professional journal limited to publication of original research reports and authoritative theoretical articles, it aims to foster the promotion and expansion of clinical psychology as an applied science. Special sections in each issue will be devoted to news, comments, and letters to the editors on matters pertinent to the profession. Communications and articles submitted for publication are welcome and should be addressed to the Editorial Office, 5 Pearl Street, Brandon, Vermont.

ORIGINAL ARTICLES are published only with the understanding that they are contributed exclusively to the *Journal of Clinical Psychology*. The publishers are not responsible for statements made or opinions expressed by contributors in articles printed in these columns. Copyrights cover publication in the *Journal of Clinical Psychology* and articles may not be reproduced without permission of the publishers.

MANUSCRIPTS must be typewritten, double-spaced with good margins, on one side of the paper only. The publishers will not be responsible for manuscripts lost in the mails or through unavoidable accident. Retain a carbon copy after sending the original to the editor. It is requested that all articles be as brief as consistent with clarity. All manuscripts are subject to editorial modification.

THE STYLE of all manuscripts including bibliographies must be in accordance with the instructions given in: Anderson, J. E., and Valentine, W. L. The preparation of articles for publication in the journals of the American Psychological Association. *Psychol. Bull.*, 1944, 41, 345-376. References should be arranged alphabetically according to the directions given above. Full address of the author should appear somewhere in the article, preferably at the end.

REPRINTS are furnished on order only and must be requested on the order blank provided when galley proofs are returned. Change of address notices should be sent one month before moving and should include old and new addresses. Undelivered copies will be remailed only at subscribers' expense.

The *Journal of Clinical Psychology* is published quarterly, appearing in January, April, July and October, at Rutland, Vermont, by Frederick C. Thorne, M.D. Subscription price $7.50; Canadian subscriptions $8.00; foreign subscriptions $8.50 including postage; single copies $2.00. Entered as second class matter January 15, 1945, at the post office at Burlington, Vermont, under the Act of March 3, 1879. Reentered at the post office at Brandon, Vermont. Additional entry at the post office at Rutland, Vermont. *Title registered, U. S. Patent Office.* Editorial and Business Offices, 5 Pearl Street, Brandon, Vermont.

*Copyright, 1955, Frederick C. Thorne, M. D.*

# THE KROUT PERSONAL PREFERENCE SCALE: A FACTOR-ANALYTIC STUDY[1]

ROSS STAGNER, EDWIN D. LAWSON, AND J. WELDON MOFFITT

*University of Illinois*

## PROBLEM

For many years psychologists have looked for objective ways of testing hypotheses derived from Freud's theories of personality. Perhaps most attention has been given to the various mechanisms of adjustment.(2) Considerable research has also been done on relationships between parents and children(4) and the stages of psychosexual development.(3) The present research attempts to apply the statistics of factor analysis to a problem arising from this theory.

The development of a pencil-and-paper test, essentially projective in character, intended to estimate the extent to which a person shows tendencies reflecting the various stages of psychosexual development described in Freudian literature, has recently been reported by Krout and Tabin(1) with reliability and validity data. The authors make it clear that they are dealing with each stage as a gradient rather than as a sharply delimited category of behavior. This avoids many of the criticisms commonly offered regarding the Freudian theory of psychosexual development.

The items incorporated in the test call for expressions of liking or disliking certain kinds of objects, activities and relationships. In a general way we can say that, if a person perceives as attractive those kinds of goal-objects which involve oral gratification, this is indicative of an oral component in his present personality structure. And if one person likes eight of ten, while another likes only five of ten such items, this presumably indicates a stronger oral aspect in the former. (Liking only two out of ten would also indicate oral influences, but they would be structured differently—see below.)

The Krout Personal Preference Scale (hereinafter referred to as KPPS) extends orthodox Freudian theory somewhat in that it specifies ten stages or patterns of psychosexual development. (The term "stages" is somewhat inappropriate for some of these, since it implies a chronology which often cannot be substantiated.) These ten patterns are described in the Manual[2] as follows:

I. PRENATAL. Items symbolic of prenatal comforts. "It is assumed that an individual who rates low on these items feels secure and thus tends to be active, outgoing, and inclined to seek new experiences; while one who scores high is insecure and tends to be regressive and inclined to cling to the past."

II. EARLY ORAL. "One who is low here may be considered taciturn, pessimistic, inclined to withdrawal from contact; while one who is high tends to be communicative, optimistic, inclined to seek close object relations."

III. LATE ORAL. "A low score here should represent adequate control of verbal hostility—hence, altruistic, tolerant social behavior; a high score, on the other hand, should point to an oral-sadistic type of person who drains his hostility via oral channels—hence, tends to be subtly punitive (sarcastic) in his social attitudes."

IV. EARLY ANAL. "A high score should represent an emotionally labile, 'easy going' sort of attitude, while a low score, on the other hand, should spell a tendency to emotional rigidity."

V. LATE ANAL. "Here a low score indicates irregular, rather inconsistent behavior trends. A high score points to a methodical, punctilious, stubborn, parsimonious type of individual."

VI. NARCISSISTIC: "Here a low score should imply a self-effacing type of attitude, and should point to a relatively low aspiration level. A high score, on the other hand, should imply ambition. However, as ambition is used here, it refers to aspiration rather than drive."

VII. FEMININE. "A high score on this section should point to a submissive—'cooperative' attitude (in a male or female), because the traits are all of a passive-receptive nature."

---

[1]We acknowledge with thanks a financial grant-in-aid from the Social Science Research Council to Dr. Krout, which made possible a substantial part of the statistical work here reported.

[2]*A Guide to Personal Preference Scale*, copyright 1953, by Maurice H. Krout and Johanna Krout. Available from Chicago Psychological Institute, 185 N. Wabash, Chicago 1, Illinois.

VIII. MASCULINE. "Here a low score should indicate a generally unaggressive, apprehensive type of personality, in a male, and perhaps a normal attitude in a female; whereas a high score should point to a masculinoid-aggressive make-up, in a male, and perhaps an overcompensated masculinity in a female."

IX. INTRA-FAMILIAL SUBLIMATION. "A low score here denotes an insensitive, unsentimental, unsublimated type of attitude and relatively loose personal ties. A high score points to sublimated, sensitive, emotionally mature attitudes, high personal loyalty, acceptance of authority, and enjoyment of personal services."

X. SOCIAL SUBLIMATION. "A low score stamps the testee as a socially immature individual, unconcerned about impersonal situations, unable to conceptualize his relationships; whereas a high score should stamp him as a socially mature personality with strong feelings about the larger world, and a tendency to leadership."

Ten items for each of these ten patterns were chosen to make up the test. Thus there is a total of 100 items, 10 in each subtest. Each item represents some object, activity or relationship to which the person can respond with liking, indifference or dislike. Sample items are shown for each pattern in Tables 1 and 2. These kinds of items were chosen because it was felt that they would not activate ego-defenses, and could therefore reveal hidden dynamics which would resist direct frontal exploration[3].

*The problem of functional unities.* Many of these items have a certain "face validity" in that they are patently related to oral and other mechanisms. Other items were selected from the psychoanalytic literature because it was alleged that analysis had revealed some fixation at an early developmental level as the dynamic determinant behind such behavior. The items have, therefore, a certain theoretical unity within each subtest. Empirically, however, the important question is: Do we actually find functional unity within each group of items? Is there tangible evidence that these ten sub-tests are measuring consistent patterns of behavior which are relatively independent of each other?

It seemed that the most satisfactory technique for answering this question was factor analysis. By this method it is possible to extract the most meaningful functional unities represented in a matrix of correlated items of behavior. As our results show, the method does provide a fairly satisfactory answer to the question.

## PROCEDURE

We originally planned to intercorrelate all of the 100 items in the first edition of KPPS.[4] A little exploratory work demonstrated that a matrix 100 x 100 would be impossibly bulky and that the routine clerical work would be beyond our resources. It was necessary to devise a short-cut procedure.

The solution adopted for this was to intercorrelate the items in each sub-test, giving a 10 x 10 matrix. The first centroid factor was extracted and the three items with the highest weightings (highest sum of intercorrelations with the other nine items) were chosen to represent this subtest in the larger matrix. By this method, we were able to set up a 30 x 30 matrix in which all ten subtests were equally represented.

The subjects used were 200 males and 200 females, students in introductory psychology at the University of Illinois. All tests were administered in the classroom situation. Subjects responded to items on a 5 step scale (like very much, like, feel indifferent, dislike, dislike very much), but after these had been tabulated, they were dichotomized, throwing the indifferent group to one side or the other to give the split nearest 50-50. Males and females were treated separately.

Tetrachoric correlations were computed from the Thurstone diagrams. Thus we first obtained 10 x 10 matrices for both males and females, for each subtest. When

---

[3]A number of other criteria, such as normality of distribution of responses on a 5-step scale, were also utilized. See Krout and Tabin (1954) for details.

[4]It should be noted that the 1951 revision of KPPS, now being distributed, has only 20% of the items identical with those included in the first edition which is studied here. In addition, 30% of the items were rewritten, and 50% of the items are new in the 1951 revision. This raises serious doubt as to the applicability of generalizations from this investigation to the new edition, although it is possible that these are justified.

TABLE 1. COMMON VARIANCE BY DIFFERENT METHODS OF ANALYSIS: DATA ON MALES.

| Items | Centroid $h^2$ | Principal Axes $h^2$ |
|---|---|---|
| "Prenatal" | | |
| 1. Swimming in a pool | .218 | .589 |
| 2. Sleeping in a closed room | .521 | .621 |
| 3. Swinging in a hammock | .355 | .632 |
| "Early oral" | | |
| 4. Drinking strong beverages | .274 | .539 |
| 5. Taste of warm milk | .350 | .517 |
| 6. Sucking on bones | .606 | .784 |
| "Late oral" | | |
| 7. Cracking nuts with teeth | .273 | .849 |
| 8. Biting a child in jest | .636 | .805 |
| 9. Chewing on a pencil | .274 | .387 |
| "Early anal" | | |
| 10. Painting in oil | .502 | .713 |
| 11. Tolerance for others' weaknesses | .398 | .556 |
| 12. Giving away toys, clothes, or money | .590 | .694 |
| "Late anal" | | |
| 13. Collecting stamps, etc., as a hobby | .530 | .683 |
| 14. Coming always on time | .606 | .691 |
| 15. Odor of perfume | .704 | .675 |
| "Narcissistic" | | |
| 16. Winning a race | .649 | .751 |
| 17. Appearing on a stage | .597 | .731 |
| 18. Getting a newspaper writeup | .891 | .987 |
| "Feminine" | | |
| 19. Preparing a meal for friends | .544 | .723 |
| 20. Thinking of childhood | .447 | .705 |
| 21. Being caressed | .484 | .695 |
| "Masculine" | | |
| 22. Building a bookcase | .469 | .539 |
| 23. Hunting | .669 | .741 |
| 24. Taking a trip to Antarctic | .237 | .455 |
| "Intrafamilial sublimation" | | |
| 25. Studying insects or bacteria | .814 | .830 |
| 26. Working at a hospital | .624 | .823 |
| 27. Studying psychology of people | .456 | .654 |
| "Social sublimation" | | |
| 28. Making many new acquaintances | .585 | .850 |
| 29. Solving problems for others | .533 | .660 |
| 30. Advancing oneself through friendship | .706 | .722 |

the first-order factor had been extracted, it was found that the best items were not identical in all cases. We therefore decided to continue analyzing each sex separately, hence some differences will be found between the items in Tables 1 and 2. There is, however, a very large overlap, fourteen of the thirty items chosen for the major analysis being identical. Tetrachoric correlations were then computed for the 30 x 30 matrices for males and females. Centroid analyses were prepared for the two matrices. Using Tucker's rule as to the point at which to stop extracting factors, we arrived at 10 factors in both cases.

Some time after the centroid analysis was completed, an electronic computer program was developed by Gregory and Neuhaus for applying Hotelling's principal axes method of factor analysis. We therefore decided to re-analyze the two 30 x 30

TABLE 2. COMMON VARIANCE BY DIFFERENT METHODS OF ANALYSIS: DATA FOR FEMALES

| Items | Centroid $h^2$ | Principal Axes $h^2$ |
|---|---|---|
| "Prenatal" | | |
| 1. Swimming under water | .507 | .702 |
| 2. Sleeping in a closed room | .304 | .827 |
| 3. Swinging in a hammock | .453 | .674 |
| "Early oral" | | |
| 4. Taste of cream | .390 | .866 |
| 5. Inhaling smoke | .715 | .811 |
| 6. Drinking strong beverages | .867 | .873 |
| "Late oral" | | |
| 7. Cracking nuts with teeth | .377 | .762 |
| 8. Biting a child in jest | .432 | .667 |
| 9. Chewing on a pencil | .440 | .582 |
| "Early anal" | | |
| 10. Painting in oil | .243 | .662 |
| 11. Coming late to work | .466 | .702 |
| 12. Playing with mud piles | .480 | .707 |
| "Late anal" | | |
| 13. Buying bargains | .551 | .690 |
| 14. Preserving notebooks, toys, or letters | .558 | .695 |
| 15. Odor of perfume | .620 | .644 |
| "Narcissistic" | | |
| 16. Getting a movie contract | .896 | .969 |
| 17. Posing for a picture | .414 | .483 |
| 18. Appearing on the stage | .349 | .513 |
| "Feminine" | | |
| 19. Keeping a diary | .328 | .534 |
| 20. Thinking of childhood | .431 | .726 |
| 21. Being caressed | .597 | .647 |
| "Masculine" | | |
| 22. Building a bookcase | .713 | .865 |
| 23. Hunting | .377 | .754 |
| 24. Woodcarving | .698 | .720 |
| "Intrafamilial sublimation" | | |
| 25. Teaching kindergarten | .525 | .644 |
| 26. Care of crippled children | .804 | .784 |
| 27. Working at a hospital | .460 | .636 |
| "Social sublimation" | | |
| 28. Offering ideas in group discussions | .801 | .825 |
| 29. Electioneering for a good cause | .524 | .650 |
| 30. Making discoveries | .681 | .728 |

matrices by the principal axes method.[5] This method gave us at least 12 common factors for each sex group, but because we had hypothesized only ten factors, we decided to eliminate the last two.

The unrotated factor loadings, although giving some tentative support to the hypothesis that ten meaningful clusters or patterns were involved, could not be interpreted as clearly identifiable with the "stages" of development incorporated in the original instrument. Rotation seemed necessary. Fortunately, Neuhaus and Wrig-

[5]We wish to express special thanks to Dr. Wrigley for his very helpful assistance on this matter of method, and to the Graduate Research Board, University of Illinois, for financial assistance enabling us to make use of the digital computer. Leonard Staugus aided us on IBM problems.

ley[5] developed their Quartimax method of rotation, using the digital computer, and by this method we were enabled to rotate all four matrices (centroid and principal axes, for males and females) to a pattern which approximates simple structure. (The number of zero loadings is maximized, and also the number of extreme loadings). It is these rotated matrices with which we shall chiefly be concerned.

## Results

Our main question is: Do real functional unities exist in the data? Operationally this was translated to mean: Is there a significantly high degree of interrelationship within each subtest, and relative independence between subtests?

Let us first consider the 10 x 10 tables correlating items within a subtest only. Such a table gives us 45 coefficients. Are these significantly more positive than negative? By Dixon and Mood's sign test, with $n = 45$, a split of 13/32 could be expected 1% of the time, and a split of 15/30 could occur by chance 5% of the time. Of the 20 matrices, 13 had 13 or fewer negative correlations, and one had 15 negative coefficients. Thus there is evidence that the signs of these coefficients are not distributed at random. However, the deviation from chance here could be a function of some other factor, such as a response set to like or dislike everything. That this possibility must be taken seriously is indicated by the fact that in our two 30 x 30 matrices, each with 435 coefficients, there are fewer negative signs than chance would indicate (103 for men, 116 for women). The same factor could account for Krout's high split-half reliabilities, which are mostly in the range .91 to .98 (one is .88, and one .89).

*The rotated factor loadings.* A systematic bias such as a response set cannot account for the rotated factor loadings. At best a response set might appear as a single factor; very possibly it is heavily represented in the first unrotated factor. In all four matrices, the first unrotated factor had an overwhelming predominance of positive signs.

The rotated loadings above .20 are given in Tables 3 and 4. Our principal interest is in the emergence of triads: the three items selected to represent each of the ten subtests. In 24 of the forty factors, a triad appeared with sufficiently high loadings that we felt no hesitancy in naming the factor for that triad. In 15 of the remaining factors, two items from a triad appeared with sufficiently high loadings to give some confidence that this name could be attached. In only one of the 40 factors was there no discernible trend for any group of items to be substantially loaded on that factor.

*How well do the analyses agree?* It is interesting to observe in Tables 3 and 4 that the five "postgenital" subtests come through rather clearly in the factorings. Subtests VIII (masculine), and IX (intra-familial sublimation) were clearly represented by triads in all four analyses. Subtests VI (narcissistic) and X (social sublimation) had clear triads in three analyses and two heavily loaded items in the remaining analysis. Subtest VII (feminine) had triads in two analyses (female subjects) but was less clear (two items out of three) in the two analyses with male subjects.

None of the "pregenital" subtests gave such clear results. The late oral items fitted best into the factors as rotated. The early and late anal subtests had approximate fittings. The early oral and "prenatal" items can be identified with factors, but only with considerable strain.[6]

It is interesting to observe that the males, by and large, gave much better fits in terms of theory than did the females. We have no way of estimating the significance of this difference, but it is fairly convincing as one works through the data. Decisive tests of such differences require measures of significance levels for factor loadings which seem not to be available at present.[7]

---

[6]Using a more complex system of allocating items to factors, and using only loadings above .30, Dr. Wrigley finds a "good fit" of data and theory for the males on factors 6, 8, 9 and 10, and "fair" fits on 1, 3, 4 and 5. For females he reports a good fit on 6, 8, 9 and 10, and fair on 2, 3, 5 and 7. This agrees closely with our conclusion above.

[7]Progress is being made on this problem as indicated by the papers of Rao[6] and Wrigley and Dickman[7].

TABLE 3. FACTOR LOADINGS AND ASSUMED PERSONALITY STAGES—MALES

| CENTROID ANALYSIS | | | PRINCIPAL AXES ANALYSIS | | |
|---|---|---|---|---|---|
| *Item Factor* | *Loading 1:1* | *"Prenatal"* | *Item Factor* | *Loading 2:1* | *"Prenatal" and "Early Oral"* |
| 1 | .38 | Swimming in a pool | 3 | .75 | Swinging in a hammock |
| 29 | −.37 | Solving problems for others | 5 | .48 | Taste of warm milk |
| 27 | −.36 | Studying psychology | 1 | .46 | Swimming in a pool |
| 3 | .34 | Swinging in a hammock | 29 | −.42 | Solving problems for others |
| 5 | .33 | Taste of warm milk | 13 | .36 | Collecting stamps, etc., as a hobby |
| | | | 2 | −.34 | Sleeping in a closed room |
| | | | 6 | .33 | Sucking on bones |
| | | | 14 | −.22 | Coming always on time |
| | | | 11 | .21 | Tolerance for others' weaknesses |
| *Factor* | *1:2* | *"Early Oral" and "Prenatal"* | *Factor* | *2:2* | *"Early Oral" and "Prenatal"* |
| 6 | .68 | Sucking on bones | 6 | .41 | Sucking on bones |
| 5 | .44 | Taste of warm milk | 10 | .37 | Painting in oil |
| 2 | .32 | Sleeping in a closed room | 1 | −.32 | Swimming in a pool |
| 7 | .28 | Cracking nuts with teeth | 27 | .29 | Studying psychology of people |
| 13 | .26 | Collecting stamps, etc., as a hobby | 23 | .28 | Hunting |
| 3 | .24 | Swinging in a hammock | 15 | .28 | Odor of perfume |
| 4 | .22 | Drinking strong beverages | 14 | .28 | Coming always on time |
| | | | 4 | −.26 | Drinking strong beverages |
| | | | 2 | −.26 | Sleeping in a closed room |
| | | | 24 | −.26 | Taking a trip to Antarctic |
| | | | 5 | .25 | Taste of warm milk |
| | | | 8 | −.22 | Biting a child in jest |
| | | | 29 | .20 | Solving problems for others |
| *Factor* | *1:3* | *"Late Oral"* | *Factor* | *2:3* | *"Late Oral" and "Early Oral"* |
| 8 | .52 | Biting a child in jest | 7 | .89 | Cracking nuts with teeth |
| 13 | −.52 | Collecting stamps, etc., as a hobby | 6 | .59 | Sucking on bones |
| 7 | .39 | Cracking nuts with teeth | 8 | .36 | Biting a child in jest |
| 22 | −.28 | Building a bookcase | 5 | .32 | Taste of warm milk |
| 9 | .23 | Chewing on a pencil | 10 | .20 | Painting in oil |
| 11 | −.22 | Tolerance for others' weaknesses | | | |
| *Factor* | *1:4* | *"Early Anal"* | *Factor* | *2:4* | *"Early Anal"* |
| 12 | .74 | Giving away toys, clothes, or money | 12 | .78 | Giving away toys, clothes, or money |
| 19 | .55 | Preparing a meal for friends | 11 | .61 | Tolerance for others' weaknesses |
| 20 | .42 | Thinking of childhood | 20 | .53 | Thinking of childhood |
| 11 | .41 | Tolerance for others' weaknesses | 19 | .52 | Preparing a meal for friends |
| 27 | .35 | Studying psychology of people | 13 | .38 | Collecting stamps, etc., as a hobby |
| 29 | .29 | Solving problems for others | 10 | .35 | Painting in oil |
| 22 | .24 | Building a bookcase | 27 | .34 | Studying psychology of people |
| 2 | −.24 | Sleeping in a closed room | 29 | .33 | Solving problems for others |
| 13 | .21 | Collecting stamps, etc., as a hobby | 17 | .28 | Appearing on a stage |
| 24 | .20 | Taking a trip to Antarctic | 22 | .23 | Building a bookcase |
| | | | 4 | .22 | Drinking strong beverages |
| | | | 6 | .21 | Sucking on bones |
| *Factor* | *1:5* | *"Late Anal"* | *Factor* | *2:5* | *"Late Anal"* |
| 14 | .67 | Coming always on time | 14 | .65 | Coming always on time |
| 2 | −.53 | Sleeping in a closed room | 8 | −.63 | Biting a child in jest |
| 8 | −.38 | Biting a child in jest | 2 | .53 | Sleeping in a closed room |
| 9 | −.36 | Chewing on a pencil | 4 | .47 | Drinking strong beverages |
| 3 | −.35 | Swinging in a hammock | 9 | −.47 | Chewing on a pencil |
| 17 | −.33 | Appearing on a stage | 13 | .47 | Collecting stamps, etc., as a hobby |
| 15 | .30 | Odor of perfume | 16 | .33 | Winning a race |
| 16 | .24 | Winning a race | 22 | .29 | Building a bookcase |
| 10 | −.24 | Painting in oil | 27 | .20 | Studying psychology of people |
| 11 | −.23 | Tolerance for others' weaknesses | | | |

TABLE 3 (*continued*). FACTOR LOADINGS AND ASSUMED PERSONALITY STAGES—MALES[a]

| *Item Factor* | *Loading 1:6* | *"Narcissistic"* | *Item Factor* | *Loading 2:6* | *"Narcissistic"* |
|---|---|---|---|---|---|
| 18 | .90 | Getting a newspaper writeup | 18 | .94 | Getting a newspaper writeup |
| 16 | .69 | Winning a race | 16 | .75 | Winning a race |
| 17 | .57 | Appearing on a stage | 17 | .69 | Appearing on a stage |
| 15 | .49 | Odor of perfume | 15 | .32 | Odor of perfume |
| 21 | .34 | Being caressed | 20 | .29 | Thinking of childhood |
| 20 | .28 | Thinking of childhood | 21 | .28 | Being caressed |
| 30 | .27 | Advancing oneself through friendship | 30 | .27 | Advancing oneself through friendship |
| 28 | .21 | Making many new acquaintances | | | |
| *Factor* | *1:7* | *"Feminine"* | *Factor* | *2:7* | *"Feminine"* |
| 21 | .42 | Being caressed | 21 | .71 | Being caressed |
| 8 | .31 | Biting a child in jest | 15 | .51 | Odor of perfume |
| 10 | −.28 | Painting in oil | 27 | −.33 | Studying psychology |
| 17 | −.25 | Appearing on a stage | 10 | −.31 | Painting in oil |
| 4 | .22 | Drinking strong beverages | 5 | .29 | Taste of warm milk |
| 26 | .22 | Working at a hospital | 4 | −.25 | Drinking strong beverages |
| 16 | −.21 | Winning a race | 13 | .25 | Collecting stamps, etc., as a hobby |
| 2 | .20 | Sleeping in a closed room | 29 | .22 | Solving problems for others |
| 15 | .20 | Odor of perfume | 18 | .21 | Getting a newspaper writeup |
| | | | 20 | .21 | Thinking of childhood |
| | | | 8 | .20 | Biting a child in jest |
| *Factor* | *1:8* | *"Masculine"* | *Factor* | *2:8* | *"Masculine"* |
| 23 | .74 | Hunting | 23 | .76 | Hunting |
| 10 | .50 | Painting in oil | 24 | .54 | Taking a trip to Antarctic |
| 22 | .44 | Building a bookcase | 10 | .49 | Painting in oil |
| 15 | .42 | Odor of perfume | 22 | .47 | Building a bookcase |
| 20 | .36 | Thinking of childhood | 20 | .47 | Thinking of childhood |
| 29 | .34 | Solving problems for others | 1 | .44 | Swimming in a pool |
| 30 | .30 | Advancing oneself through friendship | 30 | .33 | Advancing oneself through friendship |
| 24 | .27 | Taking a trip to Antarctic | 2 | .29 | Sleeping in a closed room |
| 19 | .22 | Preparing a meal for friends | 19 | .28 | Preparing a meal for friends |
| 13 | .22 | Collecting stamps, etc., for a hobby | 29 | .26 | Solving problems for others |
| | | | 16 | .24 | Winning a race |
| | | | 15 | .24 | Odor of perfume |
| *Factor* | *1:9* | *"Intra-familial sublimation"* | *Factor* | *2:9* | *"Intra-familial sublimation"* |
| 25 | .87 | Studying insects or bacteria | 26 | .89 | Working at a hospital |
| 26 | .74 | Working at a hospital | 25 | .84 | Studying insects or bacteria |
| 27 | .38 | Studying psychology of people | 27 | .43 | Studying psychology of people |
| 15 | −.30 | Odor of perfume | 22 | .32 | Building a bookcase |
| 24 | .27 | Taking a trip to Antarctic | 13 | .31 | Collecting stamps, etc., as a hobby |
| 4 | −.26 | Drinking strong beverages | 15 | −.30 | Odor of perfume |
| 6 | .26 | Sucking on bones | 4 | .27 | Drinking strong beverages |
| 22 | .25 | Building a bookcase | 24 | .27 | Taking a trip to Antarctic |
| 13 | .20 | Collecting stamps, etc., as a hobby | 19 | .24 | Preparing a meal for friends |
| | | | 8 | .21 | Biting a child in jest |
| | | | 11 | .21 | Tolerance for others' weaknesses |
| | | | 6 | .20 | Sucking on bones |
| *Factor* | *1:10* | *"Social Sublimation"* | *Factor* | *2:10* | *"Social Sublimation"* |
| 30 | .71 | Advancing oneself through friendship | 28 | .89 | Making many new acquaintances |
| 28 | .68 | Making many new acquaintances | 30 | .68 | Advancing oneself through friendship |
| 29 | .35 | Solving problems for others | 19 | .46 | Preparing a meal for friends |
| 19 | .31 | Preparing a meal for friends | 29 | .43 | Solving problems for others |
| 21 | .29 | Being caressed | 8 | −.30 | Biting a child in jest |
| 23 | .25 | Hunting | 21 | .29 | Being caressed |
| 11 | −.24 | Tolerance for others' weaknesses | 27 | .24 | Studying psychology of people |
| 8 | −.23 | Biting a child in jest | 17 | .22 | Appearing on a stage |
| 17 | .21 | Appearing on a stage | 6 | .20 | Sucking on bones |

TABLE 4. FACTOR LOADINGS AND ASSUMED PERSONALITY STAGES—FEMALES

| CENTROID ANALYSIS | | | PRINCIPAL AXES ANALYSIS | | |
|---|---|---|---|---|---|
| *Item Factor* | *Loading 3:1* | *"Early Oral"* | *Item Factor* | *Loading 4:1* | *"Prenatal"* |
| 4 | .50 | Taste of cream | 2 | .70 | Sleeping in a closed room |
| 5 | .36 | Inhaling smoke | 12 | .46 | Playing with mud piles |
| 1 | .28 | Swimming under water | 10 | −.39 | Painting in oil |
| 11 | −.24 | Coming late to work | 11 | −.30 | Coming late to work |
| 17 | .23 | Posing for a picture | 5 | −.28 | Inhaling smoke |
| 12 | −.21 | Playing with mud piles | 3 | .25 | Swinging in a hammock |
| | | | 27 | .22 | Working at a hospital |
| | | | 18 | .20 | Appearing on the stage |
| *Factor* | *3:2* | *"Early Oral" and "Prenatal"* | *Factor* | *4:2* | *"Early Oral" and "Prenatal"* |
| 6 | .90 | Drinking strong beverages | 6 | .87 | Drinking strong beverages |
| 5 | .66 | Inhaling smoke | 5 | .73 | Inhaling smoke |
| 11 | .34 | Coming late to work | 11 | −.47 | Coming late to work |
| 2 | .29 | Sleeping in a closed room | 23 | .39 | Hunting |
| 21 | .27 | Being caressed | 1 | .37 | Swimming under water |
| 23 | .26 | Hunting | 2 | .31 | Sleeping in a closed room |
| 1 | .26 | Swimming under water | 21 | .30 | Being caressed |
| 12 | .26 | Playing with mud piles | 15 | .25 | Odor of perfume |
| | | | 9 | .23 | Chewing on a pencil |
| | | | 12 | .22 | Playing with mud piles |
| | | | 25 | −.21 | Teaching kindergarten |
| *Factor* | *3:3* | *"Late Oral"* | *Factor* | *4:3* | *"Late Oral"* |
| 8 | .76 | Biting a child in jest | 7 | .85 | Cracking nuts with teeth |
| 7 | .57 | Cracking nuts with teeth | 8 | .75 | Biting a child in jest |
| 11 | .43 | Coming late to work | 11 | −.36 | Coming late to work |
| 13 | −.24 | Buying bargains | 13 | −.35 | Buying bargains |
| 12 | .20 | Playing with mud piles | 9 | .20 | Chewing on a pencil |
| *Factor* | *3:4* | *"Late Oral" and "Early Anal"* | *Factor* | *4:4* | *"Oral" and "Anal"* |
| 9 | .44 | Chewing on a pencil | 4 | .89 | Taste of cream |
| 27 | .39 | Working at a hospital | 3 | .37 | Swinging in a hammock |
| 12 | .37 | Playing with mud piles | 23 | −.26 | Hunting |
| 8 | −.26 | Biting a child in jest | 9 | .24 | Chewing on a pencil |
| 10 | .24 | Painting in oil | 10 | −.24 | Painting in oil |
| 5 | −.23 | Inhaling smoke | 19 | −.22 | Keeping a diary |
| | | | 15 | .21 | Odor of perfume |
| | | | 5 | .20 | Inhaling smoke |
| *Factor* | *3:5* | *"Late Anal" and "Early Anal"* | *Factor* | *4:5* | *"Late Anal" and "Early Anal"* |
| 13 | .62 | Buying bargains | 13 | .65 | Buying bargains |
| 14 | .55 | Preserving notebooks, toys, or letters | 14 | .60 | Preserving notebooks, toys, or letters |
| 25 | .51 | Teaching kindergarten | 9 | .57 | Chewing on a pencil |
| 30 | .39 | Making discoveries | 12 | .46 | Playing with mud piles |
| 9 | .38 | Chewing on a pencil | 30 | .44 | Making discoveries |
| 15 | .26 | Odor of perfume | 10 | .42 | Painting in oil |
| 23 | −.24 | Hunting | 25 | .41 | Teaching kindergarten |
| 12 | .22 | Playing with mud piles | 15 | .31 | Odor of perfume |
| 10 | .21 | Painting in oil | 23 | −.27 | Hunting |
| 1 | −.21 | Swimming under water | 21 | .27 | Being caressed |
| 21 | .20 | Being caressed | 1 | −.24 | Swimming under water |
| *Factor* | *3:6* | *"Narcissistic"* | *Factor* | *4:6* | *"Narcissistic"* |
| 16 | .92 | Getting a movie contract | 16 | .96 | Getting a movie contract |
| 17 | .50 | Posing for a picture | 17 | .60 | Posing for a picture |
| 18 | .42 | Appearing on the stage | 18 | .52 | Appearing on the stage |
| 19 | .37 | Keeping a diary | 1 | .43 | Swimming under water |
| 1 | .30 | Swimming under water | 21 | .37 | Being caressed |
| 2 | .28 | Sleeping in a closed room | 19 | .36 | Keeping a diary |
| 21 | .26 | Being caressed | 2 | .35 | Sleeping in a closed room |

TABLE 4 (*continued*). FACTOR LOADINGS AND ASSUMED PERSONALITY STAGES—FEMALES

| *Item Factor* | *Loading 3:6* | *"Narcissistic"* | *Item Factor* | *Loading 4:6* | *"Narcissistic"* |
|---|---|---|---|---|---|
| 4 | −.26 | Taste of cream | 5 | .33 | Inhaling smoke |
| 14 | .21 | Preserving notebooks, toys, or letters | 23 | −.23 | Hunting |
| | | | 8 | .21 | Biting a child in jest |
| | | | 14 | .20 | Preserving notebooks, toys, or letters |
| *Factor* | *3:7* | *"Late Anal" and "Feminine"* | *Factor* | *4:7* | *"Feminine" and "Late Anal"* |
| 3 | .64 | Swinging in a hammock | 20 | .78 | Thinking of childhood |
| 15 | .63 | Odor of perfume | 3 | .61 | Swinging in a hammock |
| 21 | .60 | Being caressed | 21 | .51 | Being caressed |
| 20 | .54 | Thinking of childhood | 19 | .50 | Keeping a diary |
| 19 | .32 | Keeping a diary | 15 | .50 | Odor of perfume |
| 14 | .30 | Preserving notebooks, toys, or letters | 14 | .40 | Preserving notebooks, toys, or letters |
| 23 | .27 | Hunting | 23 | .39 | Hunting |
| 17 | .22 | Posing for a picture | 11 | .24 | Coming late to work |
| 13 | .20 | Coming late to work | 9 | −.20 | Chewing on a pencil |
| *Factor* | *3:8* | *"Masculine"* | *Factor* | *4:8* | *"Masculine"* |
| 22 | .81 | Building a bookcase | 22 | .89 | Building a bookcase |
| 24 | .76 | Woodcarving | 24 | .79 | Woodcarving |
| 23 | .30 | Hunting | 23 | .43 | Hunting |
| 12 | .29 | Playing with mud piles | 10 | .32 | Painting in oil |
| 18 | .23 | Appearing on the stage | 12 | .30 | Playing with mud piles |
| | | | 18 | .20 | Appearing on the stage |
| | | | 1 | .20 | Swimming under water |
| *Factor* | *3:9* | *"Intra-familial sublimation"* | *Factor* | *4:9* | *"Intra-familial sublimation"* |
| 26 | .84 | Care of crippled children | 26 | .84 | Care of crippled children |
| 27 | .51 | Working at a hospital | 27 | .72 | Working at a hospital |
| 25 | .46 | Teaching kindergarten | 25 | .58 | Teaching kindergarten |
| 1 | .37 | Swimming under water | 1 | .44 | Swimming under water |
| 2 | −.28 | Sleeping in a closed room | 14 | .26 | Preserving notebooks, toys, or letters |
| 14 | .21 | Preserving notebooks, toys, or letters | 2 | −.24 | Sleeping in a closed room |
| 30 | .20 | Making discoveries | 12 | .22 | Playing with mud piles |
| | | | 10 | −.20 | Painting in oil |
| | | | 18 | −.20 | Appearing on the stage |
| *Factor* | *3:10* | *"Social Sublimation"* | *Factor* | *4:10* | *"Social Sublimation"* |
| 28 | .83 | Offering ideas in group discussion | 28 | .85 | Offering ideas in group discussions |
| 29 | .65 | Electioneering for a good cause | 29 | .76 | Electioneering for a good cause |
| 30 | .64 | Making discoveries | 30 | .66 | Making discoveries |
| 15 | .27 | Odor of perfume | 15 | .35 | Odor of perfume |
| 10 | .26 | Painting in oil | 11 | .29 | Coming late to work |
| 20 | −.22 | Thinking of childhood | 13 | .28 | Buying bargains |
| | | | 18 | .25 | Appearing on the stage |
| | | | 10 | .23 | Painting in oil |
| | | | 21 | .22 | Being caressed |
| | | | 19 | −.20 | Keeping a diary |

Different items showed wide ranges in the number of factors on which they received loadings above .20. In the principal axes analysis for males, there were seven items loaded for only a single factor, whereas there were four items with loadings above .20 on six different factors. The single factor items were: swinging in a hammock; cracking nuts with teeth; chewing on a pencil; giving away toys, clothes or money; studying insects or bacteria; working at a hospital; and making many new acquaintances. Each of these, except for "chewing on a pencil", was heavily loaded on its appropriate factor and so may be characterized as a relatively "pure" item. By contrast, the four items loaded on six factors each were: sucking on bones; biting

a child in jest; studying psychology of people; and solving problems for others. Such items probably are undesirable in a test such as KPPS, because they may be answered on the basis of numerous determinants other than the one for which each is intended.

*Relations of pregenital to postgenital characteristics.* One obviously interesting problem which can be explored with data of the present variety is this: to what extent are particular pregenital and postgenital trends related?[8] Is there any consistent tendency for particular items to be heavily loaded on factors other than their theoretical locus, and do these connections have any meaningful pattern?

Since the clearest factors are social sublimation and intrafamilial sublimation, suppose we look at these. Social sublimation, for men, has loadings on sucking on bones, biting a child in jest, and tolerance for others' weaknesses (Table 3.) These suggest orality, but only in a meager way. However, for women, social sublimation has substantial loadings on several anal and no oral items (cf. Table 4.) We are led to speculate about the greater rigidity of socializing of female children as a possible correlate of this observation.

Intrafamilial sublimation for males has loadings on both oral and anal items, but these are both positive and negative, hence not very meaningful. For females the loadings again are primarily on anal items, but less consistently than in the case of social sublimation.

A highly developed masculine trend (in males), on the other hand, seems to have a number of anal correlates, with no oral items. In this case females give parallel results. A decidedly feminine trend (in males) seems to have oral associations. For females the factors were confused, so that it was difficult to identify a suitable factor for femininity of personality. On both analyses, the factor with heaviest loadings on the feminine items also had substantial anal loadings.

The narcissistic pattern in males had no apparent pregenital associates except "odor of perfume." (There is a suggestion of passivity in the data for males, although the specific "prenatal" items do not appear.) In females the narcissistic trend was associated with prenatal and early oral items.

The general trend of these observations seems to be something like this: anal characteristics seem to be associated with aggressivity in both males and females. Anal trends are also related to sublimation in females, but no consistent pattern can be detected for males.

Orality in males seems to be associated with narcissism, and with relative femininity of personality. Females show both oral and infantile passivity (prenatal items) in the narcissistic pattern. The feminine pattern as defined by the test was difficult to identify in the data for females, but seemed to have anal rather than oral linkages.

If we work from the factors designated as pregenital, the results are compatible with these comments. For males, prenatal and early oral factors have loadings on items suggesting helpfulness in maturity. The anal factors have more loadings on items suggesting drive and aggressiveness. Helpfulness does not appear in the oral-prenatal factors for females, and even the relatively clear-cut anal factors show little relation to sublimated social behavior.

*Comparisons of the two factoring methods.* Tables 1 and 2 show the variance on each item accounted for by ten factors, derived respectively for males and for females by the centroid and principal axes methods. It is apparent that the principal axes figures are larger, but this has little importance, first, because the method uses a least-squares solution and thus extracts the highest possible fraction of the variance for each successive factor; and second, because unities were inserted in the diagonal cells as estimated communalities, whereas in the centroid analysis, the highest col-

[8]Krout and Tabin (1954) report intercorrelations ranging from −.05 to .41 for the ten subtests in the first edition of KPPS. The highest coeficients, curiously enough, are within the four subtests prenatal, oral and early anal. These range from .35 to .41. The two sublimation scores correlate only .24. Some ambiguity is introduced by the fact that these correlations include both males and females, whereas our analyses treat the sexes separately.

umn correlation was inserted. More interesting is the fact that both methods have extracted a very high proportion of the variance for most of the items. This seems to lend at least indirect support to Krout's ten-factor hypothesis.

The two methods also agree in picking out the same items as significantly loaded. If we set the minimum loading at .30, on the basis of the estimate by Dr. C. F. Wrigley that the probable error of our loadings is about .15, we find for women 36 such items on the centroid analysis, of which 32 are also above .30 in the principal axes data. For men 45 centroid items are above .30, 33 agreeing in the principal axes figures. And, as noted above, we get rather similar "fits" with the theoretical predictions by the two methods.[9]

## Summary

The first edition of KPPS was administered to 200 males and 200 females, students in introductory psychology. The ten items in each subtest were intercorrelated and the three best representatives chosen. These were similar but not identical for the two sexes. The thirty items selected were inter-correlated and factor analyzed, using both centroid and principal axes solutions. Quartimax rotation led to meaningful structure for both males and females.

The evidence favors the view that there are ten factors necessary to account for most of the variance. However, they do not fit theoretical expectations in a completely satisfactory way. The assumed postgenital stages give rather good fits; the pregenital stages were cluttered and did not separate clearly. On the whole, the evidence is interpreted as confirming the Freudian hypothesis.

The factor loadings suggest that successful sublimation is based more on anal traits in females, but on a mixture of oral and anal characteristics in males. Helpfulness in males seems loaded with orality. Aggressive masculinity is loaded with anal items in both sexes. Femininity in males seems pretty definitely related to oral trends, but in females there is a suggestion that anal influences are related to femininity.

The principal axes solution for factorization seems to give slightly more meaningful results in both sexes than the centroid method, but agreement between the two methods is high. However, much remains to be done in determining significance levels for factor data. Until this is achieved, factor analysis will also have to be considered as a projective technique.

## References

1. Krout, Maurice H., and Tabin, Johanna K. Measuring personality in developmental terms. *Genet. Psychol. Monogr.* 1954, 50, 289-335.
2. Sears, Robert R. Survey of objective studies of psychoanalytic concepts. *Social Science Research Council, Bulletin #51*, 1943.
3. Blum, Gerald S. A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, 39, 3-99.
4. Stagner, Ross, and Krout, M. H. A correlational study of personality development and structure. *J. abn. soc. Psychol.*, 1940, 35, 339-355.
5. Neuhaus, Jack O., and Wrigley, Charles F. The Quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. statis. Psychol.*, (in press).
6. Rao, C. Radhakrishna. Estimation and tests of significance in factor analysis. *Psychometrika*, (in press).
7. Wrigley, Charles F., and Dickman, Kern W. Degree of variation in factor loadings from one sample to another. *Canadian Psychol. Assn., Montreal, P. Q.*, June 5, 1954.

[9]Qualified investigators can obtain copies of the original correlations and unrotated factor loadings by writing to Dr. Stagner.

# PSYCHOLOGY INTERNSHIP TRAINING: A PANEL DISCUSSION

J. L. MC CARY
*Univ. of Houston*
*Houston, Texas*

L. S. MC GAUGHRAN
*Univ. of Houston*
*Houston, Texas*

JOSEPH BREWER
*Child Guidance Clin.*
*Wichita, Kansas*

JOSEPH DAWSON
*Southeast Louisiana Hosp.*
*Mandeville, La.*

AUSTIN DES LAURIERS
*Topeka State Hosp.*
*Topeka, Kansas*

AUSTIN FOSTER
*Texas Medical School*
*Galveston, Texas*

RUTH HUBBARD
*V. A. Hosp.*
*Waco, Texas*

CARMEN MILLER
*Southwestern Med. School*
*Dallas, Texas*

ABEL OSSARIO
*St. Louis St. Hosp.*
*St. Louis, Mo.*

## INTRODUCTION

When the Psychology Department of the University of Houston was making plans for a dedication program for its new building and laboratory in May, 1954, the problem arose as to which of the most important current issues in the area of psychology might be profitably discussed at the dedication. Many issues came to mind, but one which seemed to involve a complexity of problems, chiefly because of different views held by the different institutions, was the question of internship training.

What are the specific aims of the various agencies at which the students are to intern? Certainly there is in all agencies training psychological interns the underlying, general aim of giving the student the knowledge essential to the practice of clinical psychology, a knowledge which cannot be gained from the class room and from books. The student must gain experience with people in order to be able to apply the theories, facts and techniques he has acquired in the academic setting of the university. Despite the rather adequate requirements and discussion on internship training set forth by various committees and writers[1, 2, 3, 4], specific requirements, emphasis on various aspects of the training, type of patients in the agency, relationship with allied professional groups, etc., often vary from one agency to another.

To say that all intern training agencies should have similar training facilities and requirements is certainly questionable. Indeed, much is often gained by "specialization" at a particular agency. However, in order to understand better some of the problems of various training agencies and to receive first hand information as to what the agencies have to offer the student in way of training, and, in turn, what the agencies expect of the students, it was decided to have on the dedication program a panel of psychologists who direct the psychological internship training program for outstanding agencies in the Southwest and Midwest. These panel members represent state hospitals, medical schools, a child guidance center and a Veterans Administration hospital. The moderator of the panel is Dr. Laurence S. McGaughran, a clinical psychologist from the graduate faculty, Psychology Department, University of Houston. The following is a transcription of the presentations and discussion by the panel members.

*Dr. McGaughran:* . . . We are concerned tonight with a group discussion on a very vital part of the clinical student's training—that of his internship. As most of you probably know, probably all of you, the period of training for a clinical psychologist today generally requires four years leading to the baccalaureate degree, and four years leading to a doctoral degree in the specialty, or professional field, of clinical psychology. The third of these latter four years is frequently an internship year, during which the student attempts to apply the great mass of information which he has accumulated from his didactic training, or as the saying goes, "his book-learning." The body of people who are really responsible for making sense out of the information which we as University people try to impart to the student are, I think, ably represented by the participants we have here this evening.

. . . We have, through careful selection, gathered representative clinical directors of the most important internship facilities in the Midwest and Southwest. . . . I think that you will find that there will be differences of opinion—differences of opinion in such matters as the relative amounts of emphasis to be placed on the functions of assessment, therapy, and research; differences in opinion concerning the methods and policies involving interprofessional relationships; differences in opinion in respect to

the service load which the intern is expected to carry by the facility with which he is associated. These, and a number of other topics, are apt to be fairly controversial topics in internship training. . . . .

. . . First of these speakers, alphabetically, is Dr. Joseph Brewer, Chief Psychologist at the Wichita Guidance Center, Wichita, Kansas. . . . The second is Dr. Joseph Dawson, Chief Psychologist at the Southeastern Louisiana Hospital, Mandeville, La. The third is Dr. Austin Foster, Chief Psychologist at the Texas Medical School and John Sealy Hospital, Galveston, Texas. The fourth, Dr. Ruth Hubbard, Chief Psychologist at the V. A. Hospital, Waco, Texas. The fifth is Dr. Austin Des Lauriers, Chief Psychologist at the Topeka State Hospital, Topeka, Kansas. Sixth, Dr. Carmen Miller, Chief Psychologist at the Southwestern Medical School and Parkland Hospital, Dallas, Texas. The seventh, Dr. Abel Ossario, Chief Psychologist at the St. Louis State Hospital, St. Louis, Mo. Without taking up any further time, may I introduce the first speaker, Dr. Joseph Brewer of the Wichita Guidance Center, Wichita, Kansas. Dr. Brewer.

*Dr. Brewer:* I would like to describe briefly an internship program in a community guidance center which is essentially a psychological service, in a community of about 200,000 population. The staff is composed of 7 psychologists, 2 social workers, and 4 consulting psychiatrists, and we have plans for 4 interns in psychology. This service is primarily a service for children and their parents, although some adults are seen. And the role that an intern is expected to play finally is that of a junior staff member by the time he has finished his year, carrying out the roles primarily of therapy and diagnosis. I say therapy first because the greatest part of our work is treatment of children and counseling with parents. And the major portion of the interns' time will be devoted to learning the role of therapist with children and therapist or counselor with parents. The philosophy of our internships is one that attempts to provide resources for the individual intern, and the program has varied from year to year depending upon the composition of our interns. One of the very important things that seems necessary in a community service such as this is to acquaint the interns with the role of a community agency, and the role of the various institutions, agencies, and organizations with which such an agency works. This requires the assistance of our chief social worker in attempting to get across the concept of community organizations and the role of the various community agencies in meeting the needs of the individuals in our community. Now this means that the intern must be made acquainted with the juvenile court, the school counseling program, the children's agency, the family agency, the welfare department, the community planning council, and other agencies that work very closely with us as referral sources and as combined treatment resources.

One of the important things that we try to do in the very beginning is to assess the individual's competence in the areas of diagnosis and treatment, and try to plan with him an experience that will enable him to assume a full role of a junior psychologist by the end of his internship year. This is accomplished primarily through a review with the interns of their training, experience and attitude by the total staff. During the first three or four months of the internship year, there will be on the average 2 to 3 staff meetings a week, at which the interns will take part, listening in, participating. The formal organization of their supervision and bringing them into the program, is set up something like this: They will participate in three staff meetings, one of which will be an intake staff, one of which will be either a training staff meeting or an administrative meeting, and one will be a diagnostic staff. They will have a weekly seminar with one of our consulting psychiatrists, geared primarily around the problems of psychodynamics and psychotherapy. They will have one hour a week with one of the psychiatrists individually in which they will consult with him about individual cases, individual problems relating to their work. They will have a group seminar with me, as director of the clinic, in terms of supervising and coordinating all of their activities and primarily, in the later part of the program, a concern with their research problem, and the application of their research knowledge to problems that might exist in the community guidance center. They will have a varying number of supervisory conferences with other senior staff members, depending upon the cases they are carrying and individuals with whom they will be working on the staff.

One of the important things that seems to characterize the interns that come to us is the great preponderance of knowledge and techniques that are geared primarily for work with adults. And it seems that perhaps there has not been as much opportunity to learn about the problems of children and the techniques and theoretical constructs that will be useful in dealing with children. This is an area in which we spend a good bit of time trying to bring them up in a deficiency that we have felt exists in a number of psychology departments. I do not know whether that deficiency exists in the training at Houston or not. It seems that perhaps the V. A. has to a certain extent dominated the programs in pointing toward work with adults.

The individual intern may, and it is his own choosing, may take advantage of opportunity for personal therapy. Some of them have, perhaps less than a third of them have, with some of the psychiatrists attached to our clinic or in the community. This is not felt to be an essential part of the experience of interns or the individual in clinical training.

The intern will have spent about, on the average, six hours a week in some kind of supervisory conference. I counted up for the month of January, I think it was, and it averaged about 6 hours for each individual in some kind of supervisory contact. The greatest proportion of the intern's time is spent in work with children. We have found that the individuals who have come to us have been best able to work with children as opposed to parents. We have found that there are more problems arising from assuming the role of therapist or counselor with parents than with diagnostician or therapist with children. I do know that the interns seem to find it easier to identify with the children than with the parents, and this may be something we can kick around as to what motivates that. The intern can,

ordinarily, use the specific techniques that he has learned right off in our program, we have found, and usually there will be some technique that he will want to acquire, some particular thing that he has not had an opportunity to learn, and we can provide the learning for him. This will be done in terms of small seminar programs or if, for instance, there has been just the beginning of work in Rorschach interpretation, there will be a continuing seminar with one of the psychologists in the community who serves as a consultant coming in to that seminar in which all the staff participates throughout most of the year.

The role that the intern plays is one of being integrated rapidly into tht total staff picture. I think he finds it fairly easy to make an identification with the agency as a whole because it is essentially a psychological service, directed by a psychologist, and a major number of staff being psychologists. The other professional people, social workers and psychiatrists, are seen as helpers and colleagues in working with the clientele of our center.

The one thing that seems most difficult to really motivate on the part of the intern is relating himself to the community in a more general way. They tend to be more readily acceptant of staying within the confines of the center and I think perhaps they feel greater security there. They do not seem to have many times the kind of knowledge about a community and how it functions to provide services for its individuals. In a community center you are exposed to the community in a variety of ways, fairly directly, and you need to develop techniques of relating yourself to schools, teachers, P. T. A. groups, social agencies, institutions for children, the court, the police department, and what not. You cannot operate in isolation and just within the walls of the clinic or center. This is an avenue of work that we can help individuals very definitely in, but there seems to be lacking on the part of a number of interns the kind of motivation that would lead them toward this integration with this exposing themselves to the community generally. We try to foster this and try to help the intern see that his role as a therapist, as a diagnostician, in a community center can only really be there if he is willing to relate himself to the resources outside the center and to the community generally, and to develop techniques by which he can so relate himself.

The evaluation of the intern is not a very formal one; we have not developed a formal evaluative procedure and we have felt that his evaluation should be geared in terms of the university's needs or demands and in terms of the goals that we set jointly together, when the intern comes to us.

*Dr. McGaughran:* Thank you very much, Dr. Brewer. The second speaker is Dr. Joseph Dawson, Chief Psychologist at Southeast Louisiana Hospital, Mandeville, La. Dr. Dawson.

*Dr. Dawson:* Before speaking directly about the experiences afforded in training and research at Southeast Louisiana Hospital, it would be appropriate to describe the hospital itself since in equipment, personnel and philosophy it presents a unique situation which differs considerably from the usual neuropsychiatric hospital.

The hospital has been functioning since December of 1952. From its inception it was conceived as a facility that would offer intensive treatment to acute psychiatric patients whose illness was recent in origin and whose prognoses were favorable. Concurrently, the hospital would be staffed and equipped for major investigations in behavioral science. To this end the hospital contains laboratories in physiology, pathology, psychology and biochemistry. It also has facilities for experimental animals, and a complete surgical theatre which is well equipped for studies of human brain function.

As well as the psychiatric and psychological personnel who function in the service area, the staff also includes basic scientists in biochemistry and physiology, and it is planned to add a sociologist or anthropologist. Hospital research will have a multidisciplinary orientation with the cooperation of these personnel. The hospital is affiliated for training with L. S. U. and Tulane University, and some of the staff hold teaching appointments with the graduate and medical faculties of these institutions.

The Clinical Psychology service was established in August 1953, following the recommendations of the committee on Practicum Training of the A. P. A. At the present time there are three full time staff members and two fellows in clinical psychology. Fellows in clinical psychology are appointed at three levels of training. Level I applicants must have completed two years of graduate study in an accredited university and have departmental approval for continuation to the doctorate. Level II candidates should have completed three years of graduate study, while those at Level III must have completed all requirements for the doctorate except the dissertation and the final oral. Stipends for these training positions are $2000, $2600 and $3400 per annum. Within the next six months we hope to announce a post Ph.D. program for candidates needing to acquire additional clinical or research experience. A selection committee composed of the chairman of the Departments of Psychology at Tulane and L. S. U. together with the clinical director and chief clinical psychologist are presently reviewing new applications for appointment in September.

The hospital is divided for organizational purposes into a male and female service for both white and Negro patients. Assigned to each service is one senior staff member, two residents in psychiatry, and a staff member and fellow in clinical psychology. It is planned to have the clinical fellows rotate at six month intervals between these services.

Any staff member, resident or consultant may initiate a request for psychological consultation; however, these referrals are not done for routine evaluation, nor is there a standard psychological battery administered to each patient who is referred. The psychiatric staff usually have rather specific questions such as the evaluation of a patient's assets for psychotherapy, the presence of specific defenses, or fine points in a differential diagnosis. Both in the analysis of the test results and in the form of the final report, an aim is made to achieve an analysis of the patient's personality structure and

defenses in the form of operational hypotheses to which we are unable to assign a fiducial level of probability, but which we can often verify clinically during the patient's stay in the hospital and during the follow-up period.

A typical fellow in clinical psychology, although I am sure that no such animal exists, who has finished his laboratory testing in the University setting is quickly introduced to patients. Our aim is not to make him proficient in techniques but to enable him to learn how these techniques may be employed to reveal a functioning individual. The fellow will usually become quite proficient in techniques, through the standard check-out and observational procedures inherent in the supervisory process. The clinical fellow is responsible after he has passed through his "checkouts" for the presentation of the diagnostic findings to psychiatric staff conference and is responsible for the defense of his findings. If I may be permitted a reassuring aside, this does not represent the traumatic experience that some of us perhaps have known. The fellow is recognized by his senior staff colleagues as a professional associate and interpersonal relations between staff are excellent.

There are other kinds of diagnostic experiences in which the fellow participates. Some of these include the opportunity to observe and record during amytal interviews, physiological techniques such as EEG and to observe depth interviewing by outside consultants, and other laboratory procedures. Special emphasis is given to the organization of the psychological findings into a comprehensive scientific report. The meaningful integration of these findings is no mean problem for most psychologists in training or otherwise. We prefer a report that presents dynamics rather than diagnostic categories, although if this goal is attained the diagnostic category should not be difficult to ascertain.

Supervision is usually given to the clinical fellow at regular intervals as he works through his diagnostic problem. During the latter phases of his residency he functions as a junior staff member with a decreasing amount of supervision, and with more responsibility.

Remembering both my former experiences in a hospital and University setting in which certain problems reoccur, it has seemed necessary to provide a beginning therapeutic experience which avoids as much as possible the likelihood of trauma to patient or beginning therapist. At first inspection this does not seem a difficult problem, but after observing several of my practicum students in a state of near anxiety, I began to appreciate the threatening aspects of the situation. A teacher quite naturally does not like to admit that his level of aspiration for his student is unrealistically high. It was then that I began to think in terms of an introductory therapeutic situation which would allow the student the opportunity to interact with a patient in a relatively controlled setting. We call this the study of dynamics. It is this method that we now employ although I am sure it is not original with our hospital. After conference with one of the psychiatric staff, a patient is selected whom we think will present an optimal learning situation from the student's standpoint and who will benefit from supportive treatment. We review the psychiatric and social history carefully and structure the initial situation as follows . . .

"With this patient you are not expected to do treatment, or assume responsibility for the patient. Your primary mission is to get to know this patient and his life situation, and to attempt to derive hypotheses as to the dynamics of his behavior. This will include his needs, desires, motivations, and anything else which appears to be important for the patient. You will not be directly involved in the administrative handling of the patient, although the psychiatrist will usually want to discuss with you your contacts with the patient. You will have regular conferences with your supervisor."

The protocol of the sessions is divided into two parts primarily for convenience. The first part comprises a summary account of what the patient related during the session, with any of the comments made by his therapist. The second part consists of the student's formulation of the patient's dynamics, and the recording of his own feelings toward the patient that have arisen out of the relationship. It is in this part that we encourage the clinical fellow to "stick his neck out" with hypotheses of causation, or determination of the patient's behavior. Naturally this picture changes as the relationship progresses, but it is often possible to "test" these hypothese clinically and to derive predictive indices for the patient's future behavior.

The advantages of such an approach are fairly obvious. While relieving the clinical fellow of responsibility and minimizing the possibility of over-identification with the patient, it permits him to learn the basic process of relating to a patient, observation of important factors in the patient, to develop his own ideas of behavioral motivation, and to observe in himself his reactions to the patient. It is precisely this problem which constitutes a fundamental weakness of most non-analytic training programs in psychotherapy. Most supervisory sessions actually entail working through the dynamics of the beginning therapist more so than the patient. The most efficient way it seems to me is to have the therapist in treatment also since his staff supervisor should not be involved with the students' own problems for obvious reasons.

From this introduction to dynamics which is actually supportive treatment, the fellow may progress to a more active treatment role. In connection with this may I digress for a moment to say that we have in New Orleans, as many of you know, an analytic study group which is affiliated with the Washington-Baltimore school of psychiatry. Hospital policy permits fellows in training with us sufficient time to secure therapy with a qualified analyst of this group.

Almost if not most important from the standpoint of the patient in our hospital are two different forms of group therapy. The first of these, which is like the more traditional type of small group situation, is usually composed of about 12 patients with a therapist and a recorder. There are five groups presently active. In some groups the therapist alone leads the group, while in other groups the recorder is also a therapist. To date none of our fellows has acted as therapist but when the groups are re-

organized it is expected that the more advanced fellows will be able to serve as therapists. One of the present fellows is acting as recorder for a group.

A more unique kind of group therapeutic situation with broad objectives is the ward meeting. About three months ago, the hospital together with the Louisiana Department of Institutions, sponsored a visit of Dr. Maxwell Jones, the author of "The Therapeutic Community", to the hospital. I am sure that many of you have read of Dr. Jones' experiences in what may be called group therapy for want of a better name, in his Industrial Neurosis unit at Belmont, England. Briefly, Dr. Jones and his staff at the Belmont unit set up a therapeutic community in which each member of the staff was a therapist and the patients in this setting participated in their own treatment almost 24 hours a day. To a very great degree responsibility is assumed by the patients for themselves and each other. Traditional lines of authoritarian relationships have been modified, if not entirely altered, so that staff and patients exercise a maximum degree of communication and responsibility. Results under this approach even with a very difficult group of patients (mostly psychopathic or "untreatable") far exceeded any other therapeutic regime that had been previously attempted. Following Dr. Jones' visit, and with his advice, it was decided to inaugurate such a plan at the hospital. Several group meetings were held with the staff in which Dr. Jones participated, in which the fear, anxieties and doubts of the staff were ventilated. With a good deal of trepidation on the part of the staff the program was initiated. A "responsible worker" was designated for each ward meeting, although any of the physicians, psychologists, social workers, nurses, attendants, and other personnel were free to express whatever opinions or contribute to the discussion in any way appropriate. The situation was defined to the patients that they were free to express any ideas about themselves, the institution, personnel or anything that they desired. Meetings on the very disturbed wards were planned to last about a half hour, followed by a half hour's discussion with the personnel attending the meeting. The purpose of the latter meeting was to allow the staff to discuss problems originating in the ward meeting, and work out their anxieties growing out of the situation. In one of the first ward meetings that I conducted, the patients accused an attendant of striking a patient. Patient and attendant were both present at the meeting so you can well imagine the hostilities generated in such a face-to-face situation. With disturbed patients the discussion frequently consists of delusions or hallucinations which may be acted out directly. With more integrated patients the productions are of more conscious material. On the basis of our relatively brief experience the ward meeting is proving a fruitful experience for both staff and patients, and an excellent learning situation in group dynamics.

In our practicum program we planned to provide a place for students who were interested in the clinical area, but whose major previous training had been in the field of experimental psychology, as well as for majors in the field of clinical psychology. These students would be expected to become acquainted with problems found in the design of clinical research, and to develop research skills in this area. Since Tulane University now has no clinical program, this aspect of our program meets a need at both a local and national level.

The hospital has ample funds for the purchase or construction of needed equipment for experimental investigations. Studies either projected or in progress include: the investigation of the relationship of subliminal processes to perception in neuropsychiatric patients and non-hospitalized subjects, an attempt to derive criteria for the selection of attendants, personality studies of psychiatric nursing personnel, and various studies in the field of brain injury Interdisciplinary research is encouraged and the department collaborates in the design, execution and analysis of projects in psychiatry, physiology, nursing and other disciplines.

Our students at the third level fellowship must give major emphasis to the gathering and analysis of data for the doctoral dissertation. At any level, however, fellows are encouraged to develop their ideas into a systematic research plan and a departmental research seminar is designed to facilitate this development. In this seminar the fellows and staff present their ideas for the comments and criticism of the rest of the department.

In addition to the departmental seminar the department has participated in the organization of a hospital journal club. The purpose of the journal club is to bring outside speakers on significant research topics and to encourage staff in the presentation of their own ideas. The programs are planned so that outside speakers alternate with local staff in bi-weekly meetings.

A brief summary such as this can hardly give a comprehensive picture of the many aspects of a training program. I am sure there must be questions regarding some details which I have omitted or treated incompletely.

*Dr. McGaughran:* Thank you very much, Dr. Dawson. The third speaker, or participant, is Dr. Austin Foster, Chief Psychologist at the Texas Medical School and John Sealy Hospital, Galveston, Texas. Dr. Foster.

*Dr. Foster:* The University of Texas Medical Branch has some 300 beds for active treatment of psychiatric patients and a very large outpatient service. The psychology department supplies psychological services to the Youth Development Center, the Adolescent Clinic, as well as the Department of Psychiatry. We have a very large and active psychiatric program, all ages, all types of patients, except that we do not presently handle chronic psychoses; that is to say, committed patients are not accepted for treatment. Our work is entirely active treatment.

Our program is centered around training rather than service. We are attempting to find answers to questions: "What is the best type of field training for a clinical psychologist?" As presently constituted, the program occupied two calendar years. This creates a certain amount of apprehension among the applicants but we assure them that they only have to apply for one year at a time. Thus

far, they have all stayed on of their own accord for the second year and even though it is not in the official A. P. A. schedule they have been sufficiently motivated and involved in our program that they asked to stay with us for an additional year. Our reason for doing this is very simple: in setting up an ideal program to provide a student in clinical psychology with the kind of background he should have we have found that we cannot encompass the desired material within a single year. Therefore, we have taken the amount of time we find necessary. I think it has "paid off" well. Of the four people who have completed the program and gone on to get Ph.D.'s, one is the director of a child guidance clinic, one is the assistant to Roy Schafer at Yale, one is an assistant professor in the department of psychiatry and neurology, and the fourth one is the chief psychologist in a state hospital. So I think they *do* show that there is considerable value in terms of depth and breadth of training in the full two year program. The first year of the program is rather similar to the typical internship year in that primary emphasis is placed on psychodiagnostics, with secondary emphasis on research, and psychotherapy. It differs, however, from the usual program in that we are somewhat more theoretically oriented. In the course of that year, the student participates some 12 to 15 hours a week in seminars, and other meetings, considerably more, I think, of a formal and semi-formal teaching nature than is true of most internship programs. It makes the day rather full and doesn't leave much free time, because they are also continuing their clinical work. But because we do have an interest in theory, and an interest in developing new leads, we devote a somewhat unusual amount of time to discussion of theoretical psychodynamics, and other theoretical issues. The second year of the program is designed to emphasize primarily clinical research and psychotherapy. There is a minor emphasis on psychodiagnostics. One of the particular reasons for the two year program has been that we have not seen how you can give training in long-term therapy in a short period of time. Unless students are assigned patients immediately on arriving, they can't possibly carry patients for as long as a year in the year's time, so one of the reasons that we have the two year program is to allow the students to carry some therapy cases to completion.

Of the four who have completed their work, three were able to gather dissertation material in the course of their second year, using clinical material available in the medical school setting. We encourage that to the extent that it is agreeable to the student and to the cooperating university. Our students come with quite varied backgrounds, and we attempt to treat each one as an individual: to start "where they are" and to bring them as rapidly as possible to some kind of equivalence. For that reason, our program is essentially tutorial. We have a staff with 2 full-time teaching psychologists devoted to the training of our students. We have two other psychologists who participate less extensively in the training program. The 2 full-time training psychologists devote their day almost entirely to seminars, individual conferences, and supervision of the students. Thus, our supervision is somewhat more detailed, more, as I say, of a tutorial nature than the model internships set up. The work in therapy is supervised rather variously. Dr. Goolishian and I do a considerable amount of the supervision. Some is done by Dr. McDanald who is an analyst on the staff. We all have what I suppose you might call a very tolerant neo-Freudian approach, allowing for theoretical differences with amity and good feeling and without much cultism.

The program is frankly experimental and is subject to change at any time. If we get a new idea, that we think might help in training clinical psychologists, we'll try it. We have been exceedingly fortunate in that, to my knowledge, at no time has any request for the psychology department been turned down by the University. That is to say, we have a marked degree of freedom in developing the program to suit the needs of our interns. We render a service function to the extent that it fits into our training program. We could very easily triple of quadruple the amount of work we do in a large psychiatric setting like this. What actually happens is that we take on additional responsibilities as we feel they would be useful in our training program.

Our plan at present is to limit our number of interns and residents to four at each year level. Each year we will have four new interns coming in, and then we will have four at the second year level devoting their time primarily to psychotherapy and research. In this way, with two staff members devoting full time to supervision, we can rather effectively devote ourselves to the particular needs of the particular intern. Its an expensive program; we invest actually quite a lot of money in these people and they do not pay off in the sense of providing the University with inexpensive psychological service. So far we have had wholehearted support in developing our program. It has attracted considerable interest and we have had a number of visitors down to observe the kind of program that we have. At present, we are even considering a three year program. There is again the feeling that maybe no one would be interested in a three year program but there was that objection when we started a two year program. What we're looking for, very frankly, is the kind of program in which didactic therapy for the student can become an integral part of his preparation, rather than being an extra-curricular *ad hoc* sort of thing. And to do that would require certainly a minimum of three years. Hence, for that reason, we are considering setting up in the near future a three year training program. I do not think that it would mean the discontinuance of the present program. Again, it is an experiment which we would like to try. Out of our four who have completed their work, three were able to work in some analytic training, some didactic therapy, in the course of their two years here. But the big problem is expense. We frankly have not yet worked out a way in which we can subsidize the student to the extent of some $4-5000 dollars for individual therapy, nor have we worked out some magical device whereby all students can find themselves with 4 or 5000 dollars in ready cash. However, I do think we will be able to work out a compromise in which we can set up some type of revolving fund. It is very definitely in our plans for the future to set up such a program.

We have a very large service, some 26 residents in psychiatry, some 11 or 12 psychiatrists on the staff, so we do not have much in the way of full departmental meetings. What we do have are a series of floor conferences, in which a small group of 3 or 4 psychiatric residents, psychological residents, and 1 or 2 staff people meet together each day to discuss the patients, intake, follow-up, and disposition. We have a very pleasant relationship with psychiatry, with considerable acceptance of clinical psychology as an entity, and very excellent administrative support. There has been little published on the program and we hope to publish more as we accumulate more data. And that, at present, is the way it stands at the University of Texas.

*Dr. McGaughran:* Thank you very much Dr. Foster. The next speaker is Dr. Ruth Hubbard, Chief Psychologist, Veterans Administration Hospital, Waco, Texas. Dr. Hubbard.

*Dr. Hubbard.* This will be a description of third-year experience in a Veterans Administration Neuropsychiatric Hospital. Our philosophy about internship is that it include intensive and continuing experience with patients under qualified psychological supervision, that it means also not merely seeing patients but having some participation in the ongoing treatment plan for those patients, and working in relationship to the other professional persons in the team of workers at the hospital.

The work of the third-year trainee is, of course, built on previous experience. He will have had laboratory experience in courses at the University. He will have had a year of half-time experience in the V. A., working at a level which really fits the description of clerkship because the major portion of the time in that second year is given to diagnostic work and conferences. He thus has preliminary experience with techniques, much of it oriented pretty well toward the making of a total clinical study. During the third year he continues intensive diagnostic study of patients, now including a greater variety both of techniques and of patients. One thing we hope he will learn is to shift his orientation from techniques to persons, so he will think first of the individual he is testing and be less concerned with the devices he is using. Those devices are part of his method for becoming acquainted with the patient and understanding what resources he has for living, but the devices will become pretty well routinized so he can give fuller attention to the person himself. The supervision during this year is a little less detailed and complete; that is, the trainee does more independent thinking, and his conferences with other members of the hospital professional staff are more independent. One experience in the third year at our hospital is the use of group test findings to make an assessment of the patient. We test in small groups all incoming patients at the hospital; second-year trainees help with the administration of the tests and the making of observations; then in the third year when they have gained a good deal of experience with individual testing, trainees begin, under supervision, to write up some of the group test results. In comparison with the integration of findings from individual testing, it is a good deal more difficult to comb the relatively meager data you get from tests that are given patients in a group, and get any kind of a complete and meaningful personality picture out of them. This calls on all they know, not only of test interpretation, but also of personality dynamics and theory.

The therapy experience begins in the third year, as we plan it in our hospital. The first contact with therapy has really begun earlier because the trainee's conversations with whatever patients he is testing all through the second year are oriented toward learning as much as he can about this person, becoming acquainted with him; and discussions with his supervisor will have been along the lines of "How do these test responses fit in with what you saw of the man while he was sitting there talking with you?", "How does that patient, as you are trying to test him, make you feel?" and "How does he seem to be feeling toward you?" These questions, while he is carrying on diagnostic work, do begin to help him think in terms of relationships between people and how the patient is reacting toward the trainee. As he begins therapy, of course, the rapport of diagnostic sessions becomes a deeper and continuing relationship by which the trainee's strength may be used to help the patient face and work on his problems. The attempt to do therapy brings the trainee to see more meaning in some of his test findings and in evaluative phrases he has been using. The therapeutic work illuminates the evaluative, and the evaluative work strengthens the therapeutic.

A first step in starting individual therapy is participating in the selection of a patient for his therapy. We try to find a patient that he has tested who will be appropriate for him to take in therapy, because then he can check his own diagnostic impressions as he becomes further acquainted with the patient through therapy. He learns by this selection process the many factors involved in choosing a patient for therapy: the time element for one, because he is probably not going to be able to carry him as long as a year; the strengths and weaknesses the patient shows; why he thinks this patient can be helped; why he thinks he can be the one to help this patient; and as much as possible about what he can see as goals for therapy for this patient. All of this is good practice in clinical thinking, and sometimes it takes such a study of a good many patients before he finds one that seems appropriate for him to work with.

After some experience in individual therapy a third-year trainee will participate in group therapy. We have debated somewhat whether he should start group therapy and follow with individual therapy, or take them in the reverse order. We have asked the trainees' thinking on this subject and have generally had the response that they would rather learn to react to one person first. They think that reacting to seven or eight or ten people at once would be rather complicated. The times we have started with individual therapy and followed with group have seemed to work better, so that is the way we plan it now. The trainee has a chance in the group to observe his own feelings toward members of the group and their feelings toward him, which will be different for different members of the group; to observe their interactions with each other under the shelter of his leadership, as it were, and their benefit from that interaction. Of course, as he begins to lead a group he has a share in the selection of patients

for that group. We usually choose a time for the trainee to take over a group when it is small and additional members are coming in, so he can become acquainted with them. He does not take over a ready-made situation. He builds and integrates the group as he works with it. The choice of patients for the group is made by a staff member in consultation with the trainee. The staff member interviews each prospective patient for group or individual therapy. He talks over with the patient what the experience of therapy may be like, whether the patient thinks he can profit by it, and gives some preparation for what will be expected of him as he works with the group. The trainee observes this interview through the one-way vision window, then talks with the staff member about it, so he learns from the staff member what were the things the staff member took from that interview to use as evaluative material in deciding whether the patient should come into the group or not.

After three months or so with a group of fairly active accessible patients, a trainee starts work with another sort of group. As we have worked it so far, he generally has begun to help with one of the therapy groups on the insulin ward. It is a very different sort of patient; some are very withdrawn, some are active but go through quite lethargic periods during insulin treatment. If his first experience with group therapy were with one of these groups, he would become discouraged and give up long before the group had had much chance for life. But after some experience with a fairly active group a trainee does carry on and can see interactions in this group well enough to use them for fostering group feeling and for development of individuals in the group.

We have several groups of patients on chronic wards which trainees have not yet helped with. We have been experimenting to see what could be done with these groups of chronic patients and are about ready to have trainees help with them. One of them is made up of very withdrawn, chronic patients who spend most of their time just sitting. Group therapy for many of them is the first place where they have even talked to each other on the ward. They do not know each other; they sit beside each other on the settee all day and do not know each other's name. When the psychologist came into the ward he shook hands with some who had not shaken hands with anyone nor said a word to anyone for long weeks. Group therapy there is doing well if it gets people talking to each other, knowing each other, speaking names. Then there is another group of very aggressive, active chronic patients, a very different sort of group. The men walk around in the group, sometimes walk out, are generally active and hostile; this would be disconcerting to a trainee if he had not had some group therapy experience before. There is another chronic group who talk freely, but are often irrelevant, so it is very difficult to make out what meaning they are trying to convey. So there is a good deal of variety in groups for trainees to lead when they are interested and feel secure in leading groups.

As for supervision, during the second year the diagnostic work is discussed in considerable detail with the supervisor, but during the third year in less detail. The trainee's therapy work is observed by the supervisor. Interviews are held in the room that has a one-way vision window and are also recorded so trainees may play them back. Very often trainees do play them back and become aware of their own activity in the interview as they could not have become aware of it while the interview was going on. Nor would they be likely to accept their activity if the supervisor merely reported it to them. From the recording the trainee hears himself as he actually does it and comes in to supervisory conference with a good deal of interest and surprise saying, "I sure was going after him, wasn't I!" He learns things about his own behavior that he could not know in any way other than via recordings. In the beginning of his therapy experience, an hour or more of supervisory discussion may be needed per individual or group session; later there may be only one conference per week on individual or group therapy. For some therapy groups we have co-therapists, two leaders equal in leadership responsibility conducting a group together. Co-therapists may be a trainee and a staff psychologist, psychologist and psychiatrist, social worker and psychiatrist, or other combinations of professional people. This has certain values for the group and it surely has values for the trainee who shares a group with another professional person. The supervisor's conferences with a trainee emphasize the activity of the patient and the relationship between patient and trainee rather than the trainee's personal involvement. It seems a more valuable learning experience if the situation can be so set that the trainee becomes aware for himself of the feelings he is expressing with the patient, rather than having the supervisor be the first to call attention to them. For this reason the supervisor brings up, particularly, problems relating to the professional activity and deals with the emotional problems as they reflect themselves in this.

In the third year the trainee does participate a good deal in the planning for patients with whom he is working. He attends ward conferences and keeps in close touch with the ward psychiatrist, with the social worker, with the vocational advisor, particularly for those people he has in therapy. He is able to follow their experience through the hospital, and then as they leave, to follow what happens to them outside. The value of that experience for the trainee is that he learns how this kind of personality, which he has seen through diagnostic or therapy activities, functions later in active life outside, what he can do, what he cannot do, where the difficulties are in his adjustment. The conferences also give him something that is very important as he is going out to work somewhere as a psychologist, that is, contact with the work of other personnel about the hospital. Our insulin ward, for example, has a very close-knit intensive treatment program, so when the trainee works on that ward he becomes acquainted with the contributions of all the other people who work there, aides, nurses, occupational therapists, the dance instructor, the psychiatrists. It is an excellent place for him to learn how one can interpret psychological findings meaningfully to other people and what those findings mean in terms of a patient's behavior on the ward and his relationship to other people. The trainee begins to find himself as a psychologist among other professions. Of course, the business of a psychologist is not merely giving tests; it is functioning to help people. Tests are one device, therapy is another; and their value is

greater as they are coordinated with the work of other professions and become part of integrated planning for the patient. It takes time and many conferences for the trainee to learn to take his place in this planning.

Research work that is done by trainees in our hospital is generally done during the third year. It does not appear in the second year because generally trainees are not yet secure enough in clinical techniques or clinical thinking to begin to think of relevant problems. In the fourth year they may be working on data collected during the third year. There are five dissertations for which the data have been collected at our hospital. Two of them depended entirely on patients from our hospital; in three, patients from our hospital were compared with various other sorts of groups. Each of the men who wrote these dissertations used two or three months practically full time toward the end of their third year collecting data. Collecting the data was good clinical experience, since they were working with patients, having to evaluate them in some way and make the evaluation clinically valuable, just as if they were doing the testing for hospital purposes. Gathering research material is as much a part of their clinical training as anything else they do during the third year. One dissertation is underway right now by a third-year trainee; on the insulin ward he is studying what value psychotherapy has in addition to insulin therapy. He has set up the study in a way that uses standardized procedures; psychiatrists, social workers, occupational therapists, nurses, aides, are all helping with the evaluation of these patients before, during, and after therapy. We hope he will get results showing that psychotherapy does add something to the insulin therapy but we do not know yet. Only a few experimental patients have finished the course of treatment so far. However, it is a problem that supplies good research training for him, and one in which the hospital is exceedingly interested.

*Dr. McGaughran:* Thank you, Dr. Hubbard. The next participant is Dr. Austin Des Lauriers, Chief Psychologist at the Topeka State Hospital, Topeka, Kansas. Dr. Des Lauriers.

*Dr. Des Lauriers:* For the past four years at Topeka State Hospital, we have attempted to develop a Training Program in Clinical Psychology which would serve on the one hand to crystalize for the students the considerable amount of theoretical knowledge which they brought with them to this internship and which would, on the other hand, give to the development of their professional career in psychology a more definite and stable orientation. We do not conceive of the Internship in Clinical Psychology as just an opportunity for graduate students to be exposed to a multitude of varied types of experiences within the field of applied clinical work. Rather, we developed our program having in mind the progressive emergence in the student of a clinical way of approaching their field of work and of a professional attitude that would give integrity and status to their professional careers. This meant, first of all. that the professional staff involved in the training program would approach their responsibilities towards the trainees in the context of a mutual learning situation where the supervisor is not a master with all the answers and where the student or the intern is not an inadequate and helpless individual to be spoon-fed. It meant, furthermore, that the program would have to be structured sufficiently to allow evaluation of development or progress in the intern's work, while elastic enough to encourage initiative and drive in the student.

Thus, right from the beginning of his internship the student is considered and treated as a professional colleague. He is assigned to a section of the hospital and participates to the extent of his abilities in all phases of psychological services on that section. Through this assignment the staff psychologist on this section becomes automatically his supervisor. In order to develop early in his training a global approach to problems of personality as they are reflected in the variety of test results, the intern is made responsible from the first weeks of his internship for as complete an examination of referred patients as his training and experience will allow him to do. He is required to write a detailed report of his findings, giving as accurate a description of the patient examined as he can and integrating in a consistent personality picture his test findings. This work is closely supervised, the emphasis being on how well the intern knows the patient from his examination, how clinically complete were his observations and how technically accurate were his testing procedures. This approach enables the supervisor to make an early evaluation of the intern's technical competence, of the quality of his attitude in examining patients, and of the breadth and scope of his knowledge of personality theories and problems of psychopathology.

To supplement this supervision and to some extent bridge the gap between theory and practice, we have included in our internship a series of workshops and seminars. These have been devised specifically for the interns and are conducted by personnel of the Menninger Foundation either from our hospital, from the Menninger Clinic or from Winter Veterans Administration Hospital. We firmly believe that the ultimate profit a student can derive from the multitude and variety of practical experiences in examining patients depends essentially on the sytematic frame of reference to which he can relate these experiences and by which he can organize and evaluate them. It is part of our program to offer such systematic views in personality theory and in problems of psychopathology.

To emphasize the development of sound professional attitudes and practices in the student, we assign him to a section of the hospital for four months, a long enough period to enable him to become identified with the work going on in that section and to work closely with psychiatrists, social workers and other professional personnel. He participates in all professional meetings of the section and thus becomes a member of the professional staff there. By having each section of the hospital operating on a autonomous basis and being a complete hospital by itself, we are able to offer the intern a unique opportunity for close teamwork and have a great variety of clinical experiences.

Finally I should like to say a word about opportunities for psychotherapy in our internship program. Such opportunities exist. But we do not encourage the interns in this work until the latter part

of the internship year. And even then we prefer to have them participate in therapy case conferences or in seminars in psychotherapy than have them directly involved in therapeutic work. Our position here is based on the conviction that a sound basis for doing psychotherapy is better than small doses of therapeutic experience of a haphazard quality. One year is too short to give the interns everything.

Our program is far from perfect. We are constantly trying to improve it: first, by having frequent discussions between supervisors of the problems and methods involved in productive supervision; second, by having mutual evaluations done by both the intern and his supervisor; third by remaining alert and open-minded with respect to internship programs elsewhere, such as those I have been privileged to hear outlined in this discussion this evening.

*Dr. McGaughran:* Thank you very much, Dr. Des Lauriers. The next speaker on the program is the Chief Psychologist at the St. Louis State Hospital, St. Louis, Mo. He is Dr. Abel G. Ossorio.

*Dr. Ossorio:* I feel like an echo repeating the things that you have heard five or six times tonight, but I would like to describe to you the training program in clinical psychology that we shall have during next year at our hospital. I would like first to describe the hospital itself. It is a 3600 bed, state mental hospital located in Saint Louis, and having on the same grounds an outpatient unit. About 400 patients are admitted into the hospital every year; about 600 to 800 patients are in treatment of one kind or another at any given time. The outpatient clinic has an intake of between 200 and 300 patients a year, and there patients are admitted primarily for purposes of evaluating their potentialities for psychotherapeutic treatment. The staff of the hospital consists of a psychiatric staff of eight, a number of psychiatric residents (at this time, six), a social work staff of ten people, and a psychological staff of six. The psychology staff provides services for both the outpatient clinic and the hospital, and interns are rotated in their training through both these services. The hospital itself is a training hospital for all of the psychiatric professions. It is accredited for the training of psychiatric residents, for social workers, occupational therapists, nurses, and other professional personnel. Our clinical training program is two years old. It has undergone some changes. We expect it to change still more in the future, we hope in a positive direction.

I might tell you something about the objectives which we hope to achieve through the internship. The most obvious is, of course, to help the trainee to acquire and increase his technical skills, and to increase his competence in the various activities which we feel he may be called upon to perform in his professional career. These activities are conventionally listed under the general headings of psychodiagnostic testing, psychological treatment, and research. However, we believe that the clinical psychologist may acquire other legitimate functions in a psychiatric institution. In our hospital the teaching of other psychiatric personnel is one such function. Another function is that of acting as a consultant to other departments of the hospital on such problems as personnel selection, the proper evaluation of work performance of students, and in the construction of examinations. The intern will participate in these activities, in addition to his more conventional training.

The acquisition of professional skills is, of course, important. An equally important objective is the development of what we can call a professional attitude in all of one's working relationships. In our training program we are making a systematic effort to define for the intern what this professional attitude implies in his dealings with patients and with other personnel of the hospital. We feel that it is during this year the trainee should define and develop feeling for his professional role.

Now, I might tell you a little bit about the training program itself. The first six weeks that the intern spends in the hospital is a period of assessment of the trainee, and on the trainee's part, a period of familiarization with the hospital. During these six weeks he does no formal diagnostic work. He is assigned to the various kinds of conferences, inter-disciplinary conferences, which are established both in the hospital and the clinic for purposes of diagnosis and disposition of patients. He goes to teaching staff conferences where teaching cases are staffed. He goes to ward conferences where the business of the hospital is conducted. He goes to therapy conferences where cases carried in therapy by psychiatric residents, psychologists, and social workers are presented and discussed. He receives lectures from members of all the departments of the hospital concerning the functions of each of these departments and their particular contributions in terms of services to the patient. He observes the psychological treatments, group psychotherapy, psychodrama, multiple therapy, the somatic and chemical treatments, and lobotomies. At the same time, he is given assignments in which he utilizes all of the testing techniques in his repertoire. These testing assignments and interviews with patients are recorded and gone over with supervisors. His techniques of test administration are observed, deficiencies are corrected, and a plan of training in either new techniques or concentrations in techniques where his training may be deficient is planned. All of this occurs in the first six weeks. At the same time, his diagnostic interviews with patients are recorded and gone over with the supervisor in order that the trainee may orient himself to the patient rather than to the instrument, and so that he will learn to rely more on his own particular skills of observation rather than on the instruments that he uses to observe patients.

After this period of six weeks, he then goes into the regular training period. During the first six months of the year of training, he is assigned a variety of cases designed to expose him to all of psychiatric conditions that are found in the hospital and clinic, and also to enable him to gain competence in the complete repertoire of test instruments which we feel are minimal for his functioning. Depending on his progress during these six months, supervision over administration and scoring and analysis of individual test material, of interviewing techniques, is gradually diminished, and supervision over the integration of test findings and the presentation of a comprehensive evaluation of patients is increased. Thus, although the supervisory time remains constant, the focus of the supervision shifts to

a more complex level. After the first six months, the trainee is made completely responsible for any diagnostic testing assigned to him, from the time when the case is assigned to him to the time when the diagnostic report is communicated either to a staff conference or to the individual referring the case to him.

During the second six months also, depending on his development, he begins to take cases in therapy, individual therapy. He also has an opportunity to observe psychodrama and group therapy. His interviews in individual therapy are recorded and gone over with the supervisor. At first, every interview is analyzed in this way; later, every third or fourth interview is gone over in the session.

We feel that the internship program should really extend over a period of two years. Therefore, we have set up a plan of training which would enable a student to enter either at a first year level or at a second year. The second year level of training presupposes experience equivalent to the first year that I have just described, whether gained at our hospital or in some other institution. During the second year, the training involves a continuation of the diagnostic function—in this case the trainee functions as a staff person with complete responsibility in the diagnostic area and the focus of training is on research and on psychotherapy. He gains intensive experience in carrying patients and assuming complete responsibility for the psychological treatment of patients in both the clinic and the hospital.

Now, I also would like to mention the research opportunities at the hospital which are many. The hospital has been awarded a research grant by the United States Public Health Service which will extend over a period of three years. The general objective of this research program is to develop methods for carrying out the continuing evaluation of the entire treatment program of the hospital and of the outpatient clinic. This means, in the initial phases, developing methods for the collection and codification of data for machine analysis. These methods are in the process of being established and when this is completed sometime towards the end of this year, it will be possible to initiate a great many different projects in the general area of the evaluation of treatment and the investigation of the prognostic effectiveness of various kinds of patient and diagnostic information. This means that the trainee will be able to develop a problem in these two general areas and to have at his fingertips the means for pulling out various kinds of sub samples of the population, various kinds of data which may be required for his particular research, and even to carry on this research by means of time-saving machine methods. Thank you.

*Dr. McGaughran:* Thank you Dr. Ossario. This ten minutes steals up on you before you know it, doesn't it. The final participant in this panel is the Chief Psychologist at the Southwestern Medical School and Parkland Hospital, Dallas, Texas. She is Dr. Carmen Miller.

*Dr. Miller:* At Southwestern Medical School, the internship program in clinical psychology was inaugurated in September, 1952. We had one intern that year, we have one intern this year, and we are going to have two interns starting in September, 1954. We have not merely increased the budget to allow for two internships, but we are also increasing our regular staff in psychology, psychiatry, and social work. Therefore, we will be able to maintain a ratio of at least one full-time psychologist to every intern and we will be able to provide varied opportunities for association with related professions.

Although the psychological interns are directly responsible to and supervised by the chief psychologist, they work very closely with the entire staff of the psychiatry department and with members of other departments in the medical school who call upon the psychiatry department for services. By September, in addition to the psychologists, the psychiatry department will most probably consist of one full-time psychiatrist, two half-time psychiatrists, one full-time social worker, one half-time social worker, and six graduate trainees (residents) in psychiatry. In addition, there is a large clinical faculty in psychiatry which consists of psychologists, psychiatrists and psychiatric social workers in the community who contribute some time to our program. There are also provisions in the budget for two more full-time psychiatrists and it is possible that these positions may be filled by next September. At the present time, one of our half-time psychiatrists and two of the psychiatrists on the clinical faculty are psychoanalysts.

Most of the case material comes from Parkland City-County Hospital, which is the main teaching hospital of Southwestern Medical School. Patients are referred to us from the psychiatric ward, from the psychiatric out-patient clinic, and from the other wards and clinics requesting psychiatric consultations. Patients are very largely older adolescents and adults. We have no special psychiatric in-patient facilities for children and out-patient children are referred to the Child Guidance Clinic. Occasionally, children are referred to us by the pediatrics department.

There are now two psychiatric staff conferences a week which are held at Parkland Hospital and in which the psychologist is called upon to present psychological test findings on the patient in question. There is also one psychosomatic staff conference a week which is conducted by an internist with special psychosomatic training. At those conferences, reports on the patient are also presented by the medical student, the student nurse, the social worker, and the occupational therapist. The case is then discussed in a manner which is instructive to the medical students who are present. The writing of the psychological report is usually geared to the didactic purpose of the conference. Medical students and student nurses are given orientation lectures on psychological tests and the medical student is encouraged to arrange an individual conference with the psychologist when they are working with the same patient.

Psychology is also taught in a more formal manner as part of the psychiatry curriculum. The formal teaching is usually not done by the psychological intern. However, if he has a special interest and aptitude for teaching, he does have opportunity to participate to some extent.

Ample time is also allowed for pursuing research problems. The intern is encouraged to maintain familiarity with current literature and to engage in research projects according to his interests. If he wishes, he may use our case material in his doctoral research.

At the present time, there are no formal psychology staff conferences or seminars. However, with the increase in personnel, we plan to organize regular conferences for the purpose of discussing case material, current literature, and research problems.

The emphasis in the internship program is on psychodiagnostic work. The intern is given opportunity to gain an understanding of the purpose of various tests and to become proficient in their administration and interpretation. Interpretations are made on the basis of an integration of test results with other available information about the patient. The intern receives as much help as he needs in any phase of this work and there is special emphasis on the quality of the written report. It has been our experience that it is usually several months before the intern becomes fully oriented in the program and experienced enough to feel self-confident and secure in his work. As he gains in competence, he is encouraged to assume more and more responsibility for the content of his reports. However, his work is still supervised according to his needs.

The intern is also provided opportunity to take patients for treatment. The emphasis is on learning psychodynamics and it is not expected that the intern will have had enough experience to be a qualified psychotherapist by the end of the year. Supervision in psychotherapy is provided in regular, individual conferences with a psychiatrist.

We are now operating within a framework of limited office and clinic space. However, the first of two new medical school buildings will be completed by September, 1954, and at the same time the new city-county hospital building will also be completed. Our facilities will therefore be improved considerably, and when the second medical school building is completed, the space problem will be solved.

The internship is for one year, beginning in September. At the present time, we are able to pay $2,000.

*Dr. McGaughran:* Thank you very much, Dr. Miller. I think, perhaps, if everyone is agreeable, since each of the participants has gone to considerable detail in discussing the offerings of his facility, that perhaps it would merit us less to go around again in a round robin. Perhaps we could focus problems, questions by opening up the meeting at this time to ask if there are questions of clarification or of other types from you, as a group, which you might address either the group, a volunteer, or individually. If I might permit myself the first question, just for a moment, and I'll address this to a volunteer in the group, if I were going on to an internship I think this is the question I would like my "good fairy" to tell me—no one did tell me incidentially—perhaps you would be "good fairies" to the students who haven't gone yet. What are some of the more palpable ways which the student has under his control, by which he can go wrong, or not do his best when he enters into an internship?

*Dr. Miller:* Our interns have done their best.

*Dr. McGaughran:* Of course, this probably never happened, but perhaps you have conceived of that awful time in the future when an intern will go wrong. What will he have done?

*Dr. Dawson:* Well, I think that this would indicate a defect in the supervisory process, and I think that this could be be detected fairly early. I mean I think the supervisor is responsible for getting certain insights about what the trainee needs and being able to anticipate these things as they come up. Consequently, this is why we hope we have no students that go wrong. I think another way this would be checked is to have the student at times give an impression of what he feels he needs from the training program, and to see if his own appraisal of his needs is an objective one, that is, if it checks with the supervisor. Sometimes, this isn't objective, but I had a statement from a trainee that I felt constituted a rather difficult supervisory problem that I thought might be worth reading a little later on. I think actually, if we are, or if we do our duties to the trainee in terms of supervision, then, to answer Dr. McGaughran's question, this would not come up. But many times we don't, we have to admit our short-comings.

*Dr. McGaughran:* Dr. Ossario.

*Dr. Ossario:* I'd like to make a couple of comments on that. It seems to me that it is true that where the trainee goes wrong there has been faulty supervision or faulty evaluation of the trainee by the supervisor. With respect to specific problems, I would say that the area in which the trainee is most likely to make serious mistakes is in the area of defining his relationships to other disciplines and, in one way or another, violating the structure of the social organization to which he has to adjust in the training program. And this is the kind of thing that can result in a vicious circle of interpersonal friction which then makes it extremely difficult for the intern to function effectively in the face of this interpersonal situation.

*Dr. Brewer:* One thing I might comment on is in working with parents of children, one thing that we sometimes haven't been able to judge adequately, ahead of time, is the intern's reaction to considerable hostility on the part of parents, and considerable demanding of a kind of relationship or response that the intern cannot make, and should not make. And I think it difficult to judge sometimes their reaction to considerable hostility, because you haven't put them in that situation, particularly. You have to judge that by inference, rather than from having exposed them to hostility yourself. And I think that two or three times, we have had some difficulties there which was a lack of proper evaluation on our part in assigning a particular person to the intern.

*Dr. McGaughran:* I just reminded myself, I think probably the very first person I ever saw in a clinical situation showed hostility, and I can remember my own feelings as I came out. I was highly indignant and reported to my supervisor that this person was mad at me.

*Dr. Dawson:* It bothered him if you were mad at him, too. This is why I think we follow-up this idea of having the student express his feeling in a part of the report, and we can use this to check to really see what degree of insight he really achieves with patients as we follow him along. Sometimes the hostility of students towards supervisors gets to be pretty threatening, too, and requires quite a bit of maturity on the part of the supervisor to calm ourselves and not react to it directly.

*Dr. Ossario:* It means the development of an intra-staff atmosphere in which people can be spontaneous in their expression of feelings, and this is sometimes hard to achieve.

*Dr. Hubbard:* And you think of the trainee as a person who is learning. You don't think of him as a finished person, certainly not when he comes to you, nor when he ends his third year, he isn't a finished person, maybe the rest of his life . . . most of us are still working on it. But if you think of him as going through states that everybody does in development, then you treat the things that turn up as difficulties as things to be learned through. So that it isn't that the trainee goes wrong, it is that something has happened that has caused the learning experience to take place.

*Dr. Dawson:* That's a nice way to put it.

*Dr. Ossario:* I would like to bring up a problem which I foresee if a two-year plan, such as has been discussed by myself and others here, is put into effect, and if during the second year, the thesis research is begun. This probably is not a problem where the training installation is close to the University, but I can foresee difficulties where, for example, the state hospital is 800 miles away from Houston. The problems of the supervision of the thesis, and in the design, the planning of the thesis, I feel are primarily the responsibility of the training university, rather than the training institution. Perhaps, there are some different opinions on this. I would like some comment on that.

*Dr. McGaughran:* Perhaps Dr. McCary, of our staff, would care to comment on that point from the audience.

*Dr. McCary:* I think the problems which would arise with the intern 800 miles away from his university are quite apparent—and unfortunately quite difficult to overcome. Certainly the student would not get the supervision he needs from his university staff since he would have neither the time nor money to make the very frequent trips that would be necessary. In our own case our students who are, say, in St. Louis are on a one year program, and the only students we have on a two-year internship are in Galveston with Dr. Foster. Since Galveston is only 50 miles away and the trip from the hospital to the University can be made easily in less than an hour, this is really no problem for us—especially with the close, working relationship we have with Dr. Foster and his staff.

*Dr. McGaughran:* Thank you Dr. McCary. I think we could go on here for the remainder of the evening if we allowed ourselves to do so because this is, as I said previously, a most important topic in the area of clinical psychology. However, since there are some present who have to work tomorrow, and since we can take only so much advantage of our speakers, I think it best that we close at this time. I certainly want to thank each of the speakers for the very fine contributions made here this evening, and I, for one, am looking forward to discussing these problems with you more fully during your stay with us. Thank you and good night.

## References

1. Foster, E. A., Benton, A. L., and Rabin, A. I. The internship in clinical psychology: three alternative plans. *Amer. Psychologist*, 1952, 1, 7-13.
2. Raimy, V. C. (Ed.) *Training in Clinical Psychology.* New York: Prentice-Hall, 1950.
3. Recommended graduate training program in clinical psychology. *Amer. Psychologist*, 1947, 2, 539-559.
4. Standards for practicum training in clinical psychology: tentative recommendations. *Amer. Psychologist*, 1950, 5, 594-609.

# RATER EXPERIENCE AND CASE HISTORY JUDGMENTS OF ADJUSTMENT

A. W. BENDIG

*University of Pittsburgh*

## PROBLEM

Recently several articles have discussed problems in the use of clinical judgments in diagnosis and psychotherapy. Hamlin[6] and Bordin, et al.,[2] note two variables that are immediately relevant to the problem of using judges to secure ratings of the global adjustment level of psychiatric patients from their case histories: (a) The effect of the amount or relevance of the information or context in the case histories upon judgmental reliability, and (b) the influence of the judges' previous task-relevant training or experience upon reliability.

Bendig and Sprague[1] have reported on the influence of the training and educational background of the rater on case history adjustment ratings. Rating scales with 3, 5, 7, or 9 categories were used with each scale being anchored by three descriptive phrases. Ten short (approximately 100-word) case histories culled from psychology textbooks were rated for adjustment level, using four scales differing in length, by groups of students in introductory psychology, abnormal psychology, and graduate psychology classes, and also by a group of clinical psychologists. The average interjudge reliability within the groups of introductory psychology Ss ranged from .35 to .53 (depending on scale length) with a median of .43, while the average intercorrelation among the abnormal psychology Ss ranged from .44 to .62 with a median of .50. Similar measures were .58 and .59 for two groups of graduate psychology Ss and .65 for the clinical psychologists. The correlation of mean case ratings of the clinical psychologists with the mean case ratings of the student groups was .87 for the introductory psychology Ss, .96 for the abnormal psychology Ss, and .97 for the graduate psychology Ss. The correlation of introductory psychology Ss with the clinical psychologists was significantly lower than the similar correlations for the other two groups of student raters. It was concluded that as students receive more training in psychology they agree better with each other as to what constitutes case history adjustment and also become more similar to post-PhD clinical psychologists in their judgments.

The research reported below was designed to investigate rater reliability with more extensive and longer case histories. In addition, ratings were collected from Ss differing widely in the amount of formal training in psychology. Finally, the analysis attempted to investigate not only specific types of rating errors, but also how the Ss utilized the information contained in various sections of the case history.

## PROCEDURE

*Stimuli.* The ten abstracted case histories originally selected by Cummings[3] and later used by Newton[9] were slightly edited and mimeographed along with a face-sheet of rater instructions as a booklet. From the complete case history the male S's behavior (excluding psychometric results and psychiatric diagnosis) had been abstracted under seven major case sections. The seven case sections were:

I. Appearance and interview behavior.
II. Complaints and S's evaluation of his own adjustment.
III. Occupational, service, educational, and general economic adjustment.
IV. Parental and sibling relationships.
V. Marital and/or sexual relationships.
VI. Other interpersonal relationships.
VII. Habits, interests, aims.

At the beginning of each case history the S's present age, marital status, education, and occupation were also given. The ten abstracted case histories varied in length from one to two pages of single-spaced typing.

The face-sheet of rater instructions requested the S, after carefully reading each case, to record two judgments for each case history. The first judgment was a quantitative rating of the total adjustment level inferred from the verbal and overt behavior reported in the case. The reference point for the adjustment rating was defined as the rater's knowledge of the typical behavior of the average population male of the same age, marital status, etc., as the patient in the case history. The second judgment on each case requested of the rater was which one of the seven case sections was most important or influential in determining the adjustment rating given the case by the rater, i.e., which single case section reported behavior that the rater considered most indicative of the patient's general level of adjustment.

*Scale.* The 7-point rating scale described in a previous study [1] was used with the center and two end scale categories verbally anchored with descriptive phrases and the remaining four intermediate categories left unanchored. The anchoring statements used were: Slightly maladjusted, Moderately maladjusted, and Extremely maladjusted.

*Subjects.* The total group of 61 Ss was composed of a sample of 23 undergraduate students enrolled in two daytime sections of introductory educational psychology and a second sample of 38 graduate students in psychology. The graduate psychology Ss were first to fourth year students with 23 specializing in clinical and counseling areas and the remainder in non-clinical specialties (comparative, social, etc.).

## Results

The Ss' responses from each rater group were analyzed by analysis of variance procedures and similar methods were applied to the combined groups. From the obtained mean squares estimates of group reliability, individual rater reliability, and rater bias were derived by the methods described by Lindquist [7, pp. 359-362] and Ebel [4] and previously used by Bendig and Sprague [1]. The results of these intragroup analyses can be found in Table 1. In all three analyses significant (.001 level)

Table 1. Summarized Analyses of Variance and Reliability Estimates of Case History Adjustment Ratings

| | Undergraduates (N = 23) | | | Graduates (N = 38) | | | Combined (N = 61) | | |
|---|---|---|---|---|---|---|---|---|---|
| Source of Variation | df | Mean Square | F | df | Mean Square | F | df | Mean Square | F |
| Case Histories | 9 | 98.54 | 100.55‡ | 9 | 140.58 | 203.74‡ | 9 | 237.34 | 304.28‡ |
| Subjects | 22 | 2.55 | 2.60‡ | 37 | 1.94 | 2.81‡ | 59 | 2.17 | 2.29‡ |
| Error (CHs x Ss) | 198 | .94 | | 333 | .69 | | 531 | .78 | |
| Reliability (Group) | | .990 | | | .996 | | | .997 | |
| Reliability (Individual) | | .819 | | | .842 | | | .835 | |
| Bias | | .631 | | | .644 | | | .641 | |

‡Significant at the .001 level.

case history mean squares indicated that the Ss within each rater group reliably discriminated different adjustment levels among the ten case histories. The Ss mean squares were also significant at the .001 level, demonstrating significant individual differences between Ss in the mean rating assigned by a given S to all ten cases. This latter variable has been previously termed "rater bias." In the combined group analysis the Between Groups and the Groups by Case History interaction mean squares were not significant. Individual rater reliability (average intercorrelation between the Ss) was .82 for the undergraduate Ss and .84 for the graduates. The difference of

.02 is clearly not significant. The measure of rater bias was .63 for the undergraduates and .64 for the graduates: again an insignificant difference.

Three measures of rater error were computed for each S. *Rater bias*, as previously defined[1], was the mean of the ten ratings assigned by an S to the case histories. *Rater variability* was the standard deviation of an S's ten ratings around his mean rating. *Rater psychiatric reliability* was the correlation of an S's ratings of the cases with the scale values derived from the rankings of six psychiatrists of the same case histories[3]. Each of the correlations comprising this last variable was converted to its $z$ equivalent for the statistical analysis[5, p. 409]. The mean and standard deviation of each of the three rater errors for the undergraduate and graduate Ss are given in Table 2 along with the results of tests of intergroup differences in means ($t$) and variances (F).

TABLE 2. COMPARISON OF RATER GROUPS ON THREE SOURCES OF ERROR IN CASE HISTORY ADJUSTMENT RATINGS

| | Undergraduates (N=23) | | Graduates (N=38) | | Group Differences | |
|---|---|---|---|---|---|---|
| Rater Error | Mean | S. D. | Mean | S. D. | Means (t) | Variances (F) |
| Bias | 3.93 | .62‡ | 4.02 | .44‡ | .60 | 1.97* |
| Variability | 2.15 | .22 | 1.96 | .30 | 2.79† | 1.82 |
| Reliability | 1.32 | .27 | 1.34 | .23 | .30 | 1.37 |

*Significant at the .05 level.
†Significant at the .01 level.
‡Significant at the .001 level.

These latter comparisons showed the undergraduate Ss to be significantly (.05 level) more variable in bias measures ($F = 2.79$ with 22 and 37 dF) and to have a significantly (.01 level) higher mean variability ($t = 2.79$ with 59 dF). Neither the means nor variances of the two groups were significantly different on psychiatric reliability with the mean reliability in both groups being .87. It was also possible to test the significance of intra-group variation in the three measures of rater error by F ratio and chi-square techniques. Significant intra-group variation in bias measures had already been demonstrated by the results of the analyses reported in Table 1. The chi-square test of heterogeneity of variances[5, pp. 195-197] was applied to the variability measures within each group. The resulting chi-square for the undergraduate group (4.07) was not significant ($dF = 22$), and the chi-square for the graduate group (15.53) was similarly insignificant ($dF = 37$). The intra-group variation in reliability measures was tested by applying a chi-square test[5, p. 135] to the $z$ equivalents of the psychiatric correlations within each sample. Neither the chi-square value for the undergraduate Ss (12.16 with 22 dF) or for the graduate Ss (14.61 with 37 dF) was significant at acceptable levels, indicating inter-group homogeneity in reliability.

The final analysis compared the two groups of Ss on single case histories. The mean and standard deviation of the ratings for each of the ten case histories were computed separately for the two samples of raters. From the judgments of each S as to the most important case history sections, the frequency of Ss selecting each of the seven sections was tabulated for each case. An Index of Agreement was computed from the single case frequencies in each group. Essentially the Index of Agreement[8] is a quantitative measure of the degree of concentration of the frequencies in the qualitative categories. When the frequencies are equal in all of the categories IA equals zero and when all of the responses are concentrated in one category with zero frequencies in the remaining categories IA equals one. In Table 3 are shown the mean, standard deviation, and Index of Agreement for each case history in the two groups of Ss. In addition, the case sections selected by more than one-third of the Ss in the group are given in the last two columns of Table 3. Differences between the groups in rating means, standard deviations, or Indices of Agreement were tested by

TABLE 3. ANALYSIS OF SEPARATE CASE HISTORIES FOR ADJUSTMENT BY UNDERGRADUATE (U) AND GRADUATE (G) SUBJECTS

| Variable | Mean Rating | | S. D. | | Index of Agreement | | Case Section | |
|---|---|---|---|---|---|---|---|---|
| Group | U | G | U | G | U | G | U | G |
| Case History | | | | | | | | |
| 1 | 4.96 | 4.53 | .98 | 1.03 | .36 | .39 | III | II, III |
| 2 | 6.30 | 6.11 | .98 | .82 | .56 | 1.00 | II | II |
| 3 | 1.96 | 2.66 | .85 | .89 | .19 | .38 | V | II |
| 4 | 1.09 | 1.08 | .27 | .27 | .16 | .38 | II | I |
| 5 | 2.87 | 3.32 | 1.03 | .88 | .27 | .21 | IV | III, II |
| 6 | 1.74 | 2.11 | 1.19 | .81 | .50 | .78 | II | II |
| 7 | 5.96 | 6.21 | .78 | .80 | .57 | .94 | II | II |
| 8 | 2.61 | 2.47 | 1.31 | 1.17 | .20 | .21 | IV | II, III |
| 9 | 6.13 | 5.89 | 1.19 | 1.00 | .32 | .46 | III | III |
| 10 | 5.74 | 5.84 | 1.25 | .92 | .28 | .58 | II | II |
| Rank Test of Differences (z) | .68 | | 1.93 | | 2.56* | | | |

*Significant at the .05 level.

Wilcoxon's nonparametric rank test for paired replicates[10] and the normal curve values resulting from these tests are given at the bottom of Table 3. The graduate Ss showed significantly (.05 level) larger Indices of Agreement than did the undergraduates, but the two groups did not differ significantly on case means or standard deviations. The rank-difference correlation between the case history means from the two groups was .95 and between the Indices of Agreement was .84 (both significant at the .01 level). The rank correlation between case means and IA was .67 in the undergraduate group (significant at the .05 level) and .61 in the graduate group (just short of significance at the .05 level). Case standard deviations were not related to case means or IAs in either group.

## DISCUSSION

The results of the present study suggest that interjudge reliability in rating case histories for adjustment level is a function of the length of the case history abstract. When graduate student Ss rated shorter case histories their reliability was only .58[1] compared to the .84 found here with the same scale, but with more extensive case histories. The previously found [1] difference in reliability between professionally naive and experienced raters (.15) was greatly reduced (.02) when longer case abstracts were rated. Apparently naive Ss approach experienced Ss in reliability when (a) longer rating scales are used[1], and (b) when the case abstract provides more information about the patient.

Analysis of specific rating errors showed the naive Ss to be more variable in bias measures, i.e., the tendency to over-use one or the other end of the scale, and to spread their ratings about their mean rating more than experienced raters. The undergraduate raters tended to (a) judge all of the cases as well-adjusted or all of the cases as maladjusted, or (b) dichotomize the cases, using both extremes of the rating scale and failing to use the center scale categories. However, the significantly greater frequency of these two types of errors among the undergraduate raters did not reduce their average interjudge reliability or their average correlation with psychiatric judgments of the same case histories.

The similarity of the two rater groups in reliability contrasts markedly with the dissimilarity in their use of the information contained in specific sections of the case

histories. Although each group demonstrated significantly greater than chance agreement as to which were the most important or task-relevant sections of each case, the graduate student Ss showed greater intragroup agreement on nine of the ten cases. The rater groups also differ somewhat on which specific sections were most important. In Table 3 are recorded the case sections that more than one-third of the raters within each group selected as the most important and it can be seen that only for six of the cases (case histories 1, 2, 6, 7, 9, and 10) did the two groups agree on the important sections. The graduate Ss used sections II, III, and I (in one case) exclusively, while the undergraduate Ss also felt that sections IV and V (in one case) were important. For four of the five cases in which the groups disagreed the disagreement arose because the graduate Ss continued to use sections II and III while the undergraduate Ss reported sections IV and V as most important.

These above noted intergroup discrepancies in the selection of important case sections appear related to the sizable and similar correlation within each group ($r = .67$ and $.61$) between the rank of the cases on the mean adjustment rating assigned by the Ss and the cases ranked on the magnitude of intragroup agreement on the important case sections. Apparently both naive and experienced raters derive their judgments of highly maladjusted cases from only a few case sections, while their judgments on more well adjusted cases are taken from many different case sections. Dichotomizing the cases on the basis of the mean adjustment rating assigned by both groups shows that for the five more well adjusted cases (cases 3, 4, 5, 6, and 8) the two rater groups agreed in only one case (case 6) as to the important case section. For the five more maladjusted cases the groups agreed in every instance as to the important section. For highly maladjusted cases, both naive and experienced raters use the same behavior reported in the case history on the patient's verbalizations and his occupational, educational, and economic adjustment in forming their judgments of his maladjustment level. For the more well adjusted cases the experienced rater continues to use the same area of patient behavior in judging adjustment while the more naive rater shifts to using the case patient's parental, sibling, marital, and sexual relationships in forming his adjustment judgment. In spite of this shift in the bases of his judgments the naive rater still makes just as reliable judgments, both in terms of his intercorrelation with other naive Ss and with psychiatric judgments, as does the more experienced graduate student in psychology.

## Summary

Ten case history abstracts of clinic patients were rated by groups of undergraduate ($N = 23$) and graduate ($N = 38$) students in psychology using a simple 7-point rating scale. Each rater also indicated which case history section was most important in influencing his case rating. Individual rater reliability (average interjudge correlation) was similar in both groups and the mean correlation of raters within each group with psychiatrist ratings of the same cases was the same. Comparison of the groups on specific rater errors showed the undergraduate Ss to be more variable in measures of individual rater bias and to be higher in mean intrajudge variability. Analysis of important case sections indicated that the graduate student Ss agreed better among themselves as to which were the most important sections, but that the amount of interjudge agreement on the important sections within each case was positively related to the mean maladjustment rating given the cases by both groups. Both groups of raters selected the same case sections as important for the highly maladjusted cases, but disagreed on the important sections for the better adjusted cases.

## References

1. Bendig, A. W. and Sprague, J. L. Rater experience and the reliability of case history ratings of adjustment. *J. consult. Psychol.*, 1954, 18, 207-211.
2. Bordin, E. S., Cutler, R. L., Dittmann, A. T., Harway, N. I., Raush, H. L., and Rigler, D. Measurement problems in process research on psychotherapy. *J. consult. Psychol.*, 1954, 18, 79-82.

3. CUMMINGS, S. T. The clinician as judge: the reliability and validity of judgments of adjustment inferred from Rorschach single-card performance. *J. consult. Psychol.*, 1954, 18, 243-247.
4. EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407-424.
5. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. HAMLIN, R. M. The clinician as judge: implications of a series of studies. *J. consult. Psychol.*, 1954, 18, 233-238.
7. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton, Mifflin, 1953.
8. MATTHEWS, J. AND BENDIG, A. W. The Index of Agreement: a possible criterion for measuring the outcome of group discussion. Paper read at Speech Ass., Amer., New York, December, 1953.
9. NEWTON, R. L. The clinician as judge: the reliability and validity of judgments based on total Rorschachs and as applied to evaluations of clinical case material. *J. consult. Psychol.*, 1954, 18, 248-250.
10. WILCOXON, F. Probability tables for individual comparisons by ranking methods. *Biometrics*, 1947, 3, 119-122.

---

# AN EMPIRICAL INVESTIGATION OF THERAPEUTIC "LISTENING"

GERHART R. SOMMER, BERNARR MAZO, AND GEORGE F. J. LEHNER

*University of California, Los Angeles*

## INTRODUCTION

The literature on 'listening' in psychotherapy falls into two general categories: (1) empirical studies in which certain dimensions of therapeutic 'listening' are inferred from the ability of the therapist to predict patient behavior from protocols of patient material [4, 6]; and (2) didactic articles describing "ideal" therapeutic 'listening' [3, 5, 10]. This paper explores some variables operating in therapeutic listening and presents a technique for the empirical investigation of these variables.

Most authors agree that 'listening,' as the term is applied to therapy, implies some "deeper" understanding of the patient than a mere collection of his statements at face value or a simple interpretation of the language used by him. Rogers [11] speaks of the attitude of the therapist, as listener, as an important factor in getting the patient to express himself freely. Other studies stress that the relationship between therapist and patient becomes structured through the medium of communication [7, 12]. Fiedler [2] found much agreement among therapists as to what the ideal therapy situation involves in terms of patient-therapist communication and understanding. The statement most characteristic of the ideal therapy situation was: "The therapist is able to participate completely in the patient's communication." The statement chosen as least characteristic was: "The therapist shows no understanding of the feelings the patient is trying to communicate." These two statements, at opposite ends of a continuum of desirability for the therapeutic relationship, emphasize the factor of communication.

Many of the studies infer the listening process of the therapist from the action the therapist takes vis-a-vis his patient, either in the form of a response to the patient directly, or in the form of a prediction about the patient's behavior in a projective or other life situation. The inference is made that if the therapist is able to predict the patient's behavior, then he must know something about him, and that this information has been communicated by the patient [4, 6].

The information collected by a therapist from and about the patient in 'listening' during a therapeutic interview usually remains implicit, at least during the course of the interview. These data can be made explicit, however, by analyzing the *relationship* between the therapist's verbal report of these data (usually unverbalized) and the material provided by the patient. This relationship is here defined as 'listening.'

Any conclusions about 'listening' from the responses which the therapist makes to the patient would be influenced, of course, by a second intervening variable,

namely the therapist's "philosophy of therapy." Any technique which controls this second factor would bring us closer to the 'listening' process as here defined. Reid and Snyder[9] employed a technique which approaches control of this factor. The therapists (or subjects) were given interview protocols of patients and were asked to categorize the feelings present in the material. This technique, however, structures the subjects' responses by forcing identification of feeling, whether or not the subject thinks any feeling is present. Thus, of the many hypotheses which a therapist could make about the patient, it investigates *only* those dealing with the patient's feelings. A technique exploring 'listening' during therapy should provide the widest possible latitude for the therapist to state explicitly the hypotheses he forms.

Another methodological problem in probing 'listening' involves control of the patient's stimulus material. In the present study this was accomplished by using a "tape-patient" rather than a "live-patient." This manner of presenting the patient has the possible disadvantages of eliminating visual cues, although Giedt[4] found no significant differences in ability to predict when material was presented in sound films or on a sound track only. Limiting the stimulus material to auditory presentation probably allows for greater precision in the analysis of the results.

The problem of making explicit those responses of the therapist which usually remain implicit (while he listens) involved using a technique which would not in itself determine the explicit response. For example, the question "What has the patient said?" would tend to elicit simply a restatement of the patient's verbalization, while the question "What does the patient mean?" would tend to elicit interpretive responses, regardless of whether this is the subject's usual mode of implicit response. By using a somewhat ambiguous question "What has the patient told you?" and telling the subject that he is this tape-patient's therapist, the experimental situation was structured in the direction of eliciting responses which would more closely approximate each therapist's usual mode of data-collecting or listening during therapy.

A technique which attempts to clarify the listening process of a particular therapist should, of course, be able to highlight and clarify differences in the listening process as they occur in different groups of therapists. One such difference in 'listening' may be based upon the amount of experience which a therapist has had. In this study, therefore, we have attempted to compare experienced therapists with trainees in clinical psychology.

Recent approaches to perceptual theory[1] stress the effect of an individual's expectations on his perceptions. We should expect a similar phenomenon to occur in a listening situation, particularly when the stimulus material may be interpreted in different ways. Prior knowledge about the patient is one such source of controlling expectation and is one which is often encountered in different degrees in actual clinical practice. Therefore, half of the therapists in this study were given no information about the patient, while half were given some clinical information, as described below.

## Procedure

A tape recording of an actual therapeutic interview was used as the stimulus material. The original therapist's responses and those of the patient which were direct replies to the therapist were deleted. Contextual continuity and meaning were not lost. The patient's tape was then divided into twelve units, each approximately one minute long, which constituted meaningful segments. A tone signal indicated the end of each unit. The therapist subject sat alone in an observation room which had a one-way mirror. The tape was played by the experimenters from an adjoining room and piped into the therapy room via a speaker. The subject's responses were recorded on Audograph discs. The only equipment visible to the subject was the microphone and the speaker.

The instructions to the subject, recorded on the beginning of the tape were as follows:

"This is an attempt to learn more about listening. You will hear a recording of a patient's statements in a therapy session, and will be asked to tell what the patient has told you. There are no right or wrong answers to this question and your responses will not be checked against any particular criteria. The patient's statements will be presented in approximately one-minute units. The end of the unit will be indicated by a signal, like this (signal). You should then respond to the imaginary question 'What has this patient told you?' You may take as long as you wish and say as much or as little as you like. Please talk in your normal voice, and indicate when you have finished by raising your right hand and saying 'That's all.' Remember, you are to state *not* what you would say to the patient if he were present, but merely what he has told you. In listening assume that you are this patient's therapist and that you have no previous knowledge or contact with him. Of course, you will also be unable to see the patient and you will have to rely entirely on what you hear. If you have any questions now, please raise your right hand. (Pause) Remember, you are to respond at the end of each unit to the question, 'What has the patient told you?' Please do not respond until after you have heard the signal at the end of each unit. The next voice you hear will be that of the patient."

For half of the subjects, the following statement about the patient was added to the instructions:

"The patient is a 35-year old woman who came for help after the following incident occurred. One day she went to the information desk of a large department store to return a piece of merchandise. The clerk, who was a middle-aged woman, answered her very rudely. The patient became violently angry and fainted. On regaining consciousness, she was unable to recall the incident and when her husband, who had been with her all the time, told her what had happened, she became very worried and decided to seek help."

Any questions by the subjects regarding the meaning of the question "What has the patient told you?" were answered by instructing the subject to interpret the question in his own way.

## Subjects

Subjects were nine graduate students in clinical psychology at the University of California ("Trainee" group)* and ten psychotherapists (clinical psychologists and psychiatrists) practicing in the Los Angeles area ("Experienced" group). Half of each group listened to the tape under condition I (no information about the patient) and half under condition II (with the information about the patient given above).

## Results

The recordings of each subject's responses were typed and a content analysis of the typescript was made by two of the authors in the following manner: Each typescript was "cut" into thought-units (e. g. "she is telling that she doesn't like her mother"), then each unit was assigned to one of seven categories via the method of agreement. The content categories used, empirically derived from the responses in a pilot study [13], were:

A. Statements about the patient's dynamics, attitudes, etc. (e.g., "she hates her mother").
B. Statements about persons other than the patient (e.g., "her mother was very domineering").
C. Statements about the patient's behavior in this therapy situation (e.g., "now she's showing resistance").
D. Statements about the patient's overt behavior outside the therapy situation (e.g., "she often goes out for lunch").
E. Statements about the patient's attitude toward therapy (e.g., "she thinks therapy will tear her apart").
F. Statements of uncertainty by the therapist (e.g., "I don't know what she means by this," "I didn't hear that part too well").
G. References to the pre-therapy "incident" in condition II (e.g., "that ties in with her reactions to the woman in the store").

In addition each unit was classified as "Descriptive" if it was merely a restatement or paraphrase of the patient's verbalizations or "Interpretative" if it went beyond what the patient actually said and interpreted her behavior or verbalizations.

The frequency tally by content category and for Interpretation *vs.* Description was converted into percentages for each subject and is summarized in Table 1. The

*The "Trainee" group originally included ten subjects, but the audograph discs of one subject in condition II were lost.

TABLE 1. PER CENT OF STATEMENTS IN EACH CONTENT CATEGORY BY THERAPISTS

| Therapists (By Code No.) | | | Content Categories | | | | | | | | Interpretation *vs.* Description | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | Total | I | D | Total |
| Trainees | Condition I | I | 63 | 7 | 2 | 27 | 0 | 0 | 0 | 99 | 71 | 29 | 100 |
| | | II | 58 | 7 | 18 | 9 | 1 | 6 | 0 | 100 | 62 | 38 | 100 |
| | | III | 44 | 12 | 14 | 23 | 1 | 7 | 0 | 101 | 52 | 48 | 100 |
| | | IV | 54 | 7 | 12 | 12 | 4 | 11 | 0 | 100 | 63 | 37 | 100 |
| | | V | 60 | 19 | 2 | 16 | 0 | 4 | 0 | 101 | 36 | 64 | 100 |
| | Condition II | VII | 65 | 14 | 14 | 6 | 0 | 1 | 0 | 100 | 40 | 60 | 100 |
| | | VIII | 58 | 4 | 27 | 4 | 1 | 1 | 4 | 99 | 63 | 37 | 100 |
| | | IX | 53 | 13 | 11 | 22 | 1 | 0 | 0 | 100 | 32 | 68 | 100 |
| | | XIX | Records Lost | | | | | | | | | | |
| | | XX | 45 | 17 | 18 | 12 | 0 | 9 | 0 | 101 | 40 | 60 | 100 |
| Experienced | Condition I | VI | 46 | 0 | 40 | 6 | 0 | 8 | 0 | 100 | 76 | 24 | 100 |
| | | XIV | 59 | 13 | 13 | 12 | 0 | 2 | 0 | 99 | 62 | 38 | 100 |
| | | XVI | 63 | 10 | 13 | 12 | 2 | 0 | 0 | 100 | 60 | 40 | 100 |
| | | XVII | 41 | 9 | 19 | 14 | 0 | 17 | 0 | 100 | 70 | 30 | 100 |
| | | XVIII | 41 | 13 | 28 | 9 | 2 | 6 | 0 | 99 | 53 | 47 | 100 |
| | Condition II | X | 38 | 41 | 1 | 17 | 0 | 3 | 0 | 100 | 2 | 98 | 100 |
| | | XI | 66 | 3 | 16 | 7 | 2 | 3 | 3 | 100 | 82 | 18 | 100 |
| | | XII | 54 | 9 | 14 | 5 | 3 | 11 | 3 | 99 | 90 | 10 | 100 |
| | | XIII | 58 | 8 | 10 | 11 | 3 | 4 | 6 | 100 | 84 | 16 | 100 |
| | | XIV | 52 | 16 | 8 | 11 | 0 | 12 | 0 | 99 | 36 | 64 | 100 |

differences between "Trainees" and "Experienced" subjects, and between conditions I and II, were analyzed by means of the Mann-Whitney U Test* for differences between sums of ranks[8], for each content category and for Interpretation *vs.* Description.

The difference between experienced therapists and trainees on *interpretation* was found to be statistically significant ($.01 < p < .05$), i. e., the experienced therapists did more interpreting than the trainees. This is a difference in an "expected" direction—that is, one might expect, *a priori*, that experienced therapists would tend, in listening to their patients, to go beyond what the patient explicitly *says*, whereas trainees would tend to be more cautious, and so depend more on description rather than interpretation. None of the differences between the "Trainees" and "Experienced" groups on *content* categories was significant.

## DISCUSSION

The area in which a significant difference was found—Interpretation vs Description—was the only one selected on an *a priori* theoretical basis; all the content categories were determined empirically. It may be that this latter type of content analysis, so common in mass media and public opinion analysis, is not an efficacious approach to analysis of content in therapeutic listening, and that differentiating categories are more likely to emerge from hypotheses based on "expected" (i.e.,

*The Mann-Whitney U Test determines whether the sum of ranks for one set of scores is significantly larger than for the other. In this case the following formula was used:

$$U = mn + \frac{m(m+1)}{2} - T, \qquad Cr = \frac{\frac{mn}{2} - U}{\sqrt{\frac{mn(m+n+1)}{12}}}$$

theoretical) differences. Furthermore, we expect that meaningful interpretation of the patient's behavior must consider what the patient does here and now, and so deal with the patient's defensive efforts as well as his expressions toward the therapist. It is of interest here that in both the "Experienced" and the "Trainee" groups the two therapists lowest on percentage of Interpretation also had the lowest percentages in category C, statements about the patient's behavior in this therapy situation. Perhaps a breakdown of category C into several sub-groupings, such as "Behavior toward the therapist," and "Defenses brought into play by the therapist's interpretation," etc., would help to differentiate the "Experienced" from the "Trainee" group.

It is also possible that differences might emerge if groupings of subjects were more rigorously specified. For example our "Trainees" included both first-year interns just beginning to do therapy, and individuals with as much as four years of experience in therapy. Similarly, the "Experienced" group covered a wide range of lengths and types of experience and of theoretical orientations.

Finally, we should like to suggest the usefulness of this technique as a training device for student therapists. The method provides a dynamic situation in which the student can hear and respond to a real patient; but in which he can compare his 'listening' with that of others on exactly the same material from the same patient. Also, the students and their supervisors can evaluate changes in the student's listening as a function of training. Furthermore, the use of this method allows the student to try his therapeutic wings without having to face the frequently anxiety-producing necessity of doing something to help the patient immediately; and avoids the problems involved when a disturbed person happens to be a student's very first subject in a therapy situation.

## Summary

A rationale for the investigation of 'listening' in therapy was presented. 'Listening' was defined as the relationship between the stimulus input of the patient and the implicit hypotheses of the therapist about the patient which are made explicit in the therapist's verbal response under experimental conditions. A technique for making these hypotheses explicit was presented. These responses were content analyzed and groups of subjects were compared. Experienced therapists were found to make significantly more interpretative responses than trainees, but no differences were found between these two groups on any of the seven content categories used. Analysis of the possible effect of prior clinical information about the patient yielded no significant differences. The use of the technique as a therapy training device was discussed.

## References

1. Blake, R. H., and Ramsey, G. V. *Perception: An Approach to Personality.* Ronald Press, Inc., New York, 1951.
2. Fiedler, F. The concept of an ideal therapeutic relationship. *J. Consult. Psychol.*, 1950, *14*, 239-245.
3. Fromm-Reichman, F. *Principles of Intensive Psychotherapy.* University of Chicago Press, Chicago, Ill., 1950.
4. Giedt, F. Harold. Judgment of personality characteristics from brief interviews. Unpublished Ph.D. dissertation, University of California, Los Angeles, 1951.
5. Johnson, W., and Hayakawa, S. I. Two views on listening. ETC, 1949, *7*, 3-17.
6. Luft, J. Some relationships between clinical specialization and the understanding and prediction of an individual's behavior. Unpublished Ph.D. dissertation, University of California, Los Angeles, 1949.
7. Macfarlane, T. Empathic understanding in an interpersonal interview situation. Unpublished Ph.D. dissertation, University of California, Los Angeles, 1952.
8. Mann, H. B., and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, 1947, *18*, 50-60.
9. Reid, Dorothy K., and Snyder, W. U. Experiment on "recognition" of feelings in non-directive psychotherapy. *J. Clin. Psychol.*, 1947, *3*, 128-135.
10. Reik, T. *Listening With the Third Ear.* Farrar, Straus & Co., New York, 1949.
11. Rogers, C. R. *Client-Centered Therapy.* Houghton-Mifflin, Boston, 1951.
12. Saper, B. A multi-dimensional analysis for the study of the psychotherapeutic interview. Unpublished Ph.D. dissertation, University of California, Los Angeles, 1951.
13. Sommer, J. R., Mazo, B., and Lehner, G. F. J. The dynamics of listening. *Amer. Psychol.*, 1952, *7*, 534.

# ESTIMATING IMPAIRMENT ON WECHSLER BELLEVUE SUBTESTS

C. V. JACKSON

*National Hospital for Nervous Diseases*
*Queen Square, London*

## PROBLEM

While the Wechsler Bellevue scale[30] has been shown to be of limited value in psychiatric diagnosis,[32, 33] it remains a useful starting point in examining for intellectual deficits, and indirectly for cerebral lesions, although it is not always sufficiently sensitive. The full scale technique is time consuming, and Marks[20] maintained that many subtests were repetitive measures of mostly the same variables, since inter-subtest correlations were all high. Although the several suggested methods of diagnosing organic dysfunction by means of overall deterioration indices based on weighted scores have been shown to give results little better than chance [4, 12, 14, 28, 32], it is nevertheless often helpful to ascertain whether a patient presenting with a question of cerebral lesion has impaired scores on a verbal subtest (Similarities) or on a more spatial subtest (Block Design), and some implications of such impairment have been put forward in recent literature[1, 16, 17, 24], and in studies of prefrontal lobotomy[18, 21, 23, 25].

A simple scatter-analysis procedure for detecting impairment on individual subtests was suggested by Barnett[2], although his work was unknown to the author during two years' use of the method described below. His paper mentioned tables for every age group separately, but these were not published. Also Barnett's figures did not incorporate the findings of Fox and Birren[7] for the age group 60-69, and were not smoothed to overcome the sampling errors in Wechsler's figures which are considerable, nor did he use the refinement of "decimal weighted scores" suggested below.

Any technique of scatter analysis requires a reliable baseline, or "hold score". Ideally, this should be derived from both Information and Vocabulary scores, since these have been found to be least influenced by age[6, 8, 15], by practice[5, 13], and even by such a local cerebral lesion as prefrontal lobotomy[25], provided that the cut is sufficiently anterior[29].

If only one of these scores were available, that of the Vocabulary would be preferable since it (1) is less affected by age[6, 8, 31], (2) is more sensitive in demonstrating loss on "don't hold" subtests[10], (3) has higher test-retest reliability with normal subjects[5] and various groups of patients[9, 10, 13], and higher split-half reliability[3], and (4) shows less scatter about the line of best fit representing decline of mean score with age (see below).

Rapaport[27] suggested several base-lines for scatter analysis, but finally chose the score on Vocabulary as most suitable. On the other hand, in our experience Vocabulary scores are impaired in patients with local lesions of the dominant hemisphere or of the subordinate temporal lobe, although a score "by recognition" from a four choice list of meanings such as that prepared by Glik[11] is less susceptible. However, Wechsler's monograph did not make available the means and standard deviations obtained on Vocabulary by the various age groups of his standardizing population, although he has kindly sent a "Contrived Table of Means and Standard Deviations" based on his later experience with the test (see Table 4). With this exception, the "hold" baseline is therefore confined to the score on Information[1].

## METHOD

The raw scores on all subtests are converted to "decimal weighted scores" (shown in Table 1)[2]. These may seem an over-refinement in a measuring instrument known to be imprecise[e. g. 22], but are necessary because

[1]British Psychologists are faced with the further difficulty that no adaptation of the Information subtest has yet been sufficiently standardized for use in this country.

[2]Decimal weighted scores were determined by reading off the line of best fit after raw scores had been plotted against weighted scores for every subtest separately.

TABLE 1. WECHSLER-BELLEVUE FORM I DECIMAL WEIGHTED SCORES

| SCORES | VOCAB. | P. COMP. | COMP. | D. SPAN | ARITH. | SIMIL. | P. ARR. | BLOCKS | D. SYM. |
|---|---|---|---|---|---|---|---|---|---|
| 56 | | | | | | | | | 13.4 |
| 52 | | | | | | | | | 12.5 |
| 48 | | | | | | | | | 11.6 |
| 44 | | | | | | | | | 10.6 |
| 40 | 17.0 | | | | | | | | 9.6 |
| 39 | 16.6 | | | | | | | | 9.3 |
| 38 | 16.2 | | | | | | | 16.9 | 9.0 |
| 37 | 15.8 | | | | | | | 16.4 | 8.8 |
| 36 | 15.4 | | | | | | | 16.0 | 8.6 |
| 35 | 15.0 | | | | | | | 15.6 | 8.3 |
| 34 | 14.6 | | | | | | | 15.2 | 8.0 |
| 33 | 14.1 | | | | | | | 14.8 | 7.8 |
| 32 | 13.7 | | | | | | | 14.4 | 7.6 |
| 31 | 13.2 | | | | | | | 14.0 | 7.3 |
| 30 | 12.8 | | | | | | | 13.6 | 7.0 |
| 29 | 12.4 | | | | | | | 13.2 | 6.8 |
| 28 | 12.0 | | | | | | | 12.8 | 6.6 |
| 27 | 11.6 | | | | | | | 12.4 | 6.3 |
| 26 | 11.3 | | | | | | | 12.0 | 6.1 |
| 25 | 10.8 | | | | | | | 11.5 | 5.9 |
| 24 | 10.4 | | | | | 18.7 | | 11.1 | 5.7 |
| 23 | 10.0 | | | | | 18.0 | | 10.7 | 5.4 |
| 22 | 9.6 | | | | | 17.2 | | 10.3 | 5.1 |
| 21 | 9.1 | | | | | 16.5 | 17.6 | 9.9 | 4.9 |
| 20 | 8.7 | | 18.0 | 21.2 | | 15.7 | 16.9 | 9.5 | 4.7 |
| 19 | 8.2 | | 17.1 | 19.8 | | 15.0 | 16.1 | 9.1 | 4.4 |
| 18 | 7.8 | | 16.2 | 18.3 | | 14.3 | 15.3 | 8.7 | 4.1 |
| 17 | 7.4 | | 15.3 | 17.0 | | 13.6 | 14.5 | 8.3 | 3.9 |
| 16 | 7.0 | | 14.4 | 15.7 | | 12.8 | 13.7 | 7.8 | 3.7 |
| 15 | 6.6 | 15.4 | 13.5 | 14.2 | | 12.0 | 12.9 | 7.4 | 3.4 |
| 14 | 6.2 | 14.2 | 12.6 | 12.8 | 18.7 | 11.3 | 12.1 | 7.0 | 3.1 |
| 13 | 5.8 | 13.0 | 11.7 | 11.5 | 17.3 | 10.6 | 11.4 | 6.6 | 2.9 |
| 12 | 5.4 | 11.8 | 10.8 | 10.0 | 15.8 | 9.8 | 10.6 | 6.2 | 2.7 |
| 11 | 5.0 | 10.5 | 9.9 | 8.7 | 14.4 | 9.0 | 9.8 | 5.8 | 2.4 |
| 10 | 4.6 | 9.2 | 9.0 | 7.2 | 13.0 | 8.3 | 9.0 | 5.3 | 2.2 |
| 9 | 4.2 | 8.0 | 8.1 | 5.8 | 11.5 | 7.6 | 8.2 | 4.9 | 2.0 |
| 8 | 3.8 | 6.9 | 7.2 | 4.4 | 10.0 | 6.8 | 7.4 | 4.5 | 1.8 |
| 7 | 3.4 | 5.6 | 6.3 | 3.0 | 8.6 | 6.1 | 6.6 | 4.1 | 1.5 |
| 6 | 3.0 | 4.4 | 5.4 | 1.6 | 7.2 | 5.4 | 5.8 | 3.7 | 1.2 |
| 5 | 2.6 | 3.2 | 4.5 | 0.2 | 5.7 | 4.7 | 5.0 | 3.3 | 1.0 |
| 4 | 2.2 | 1.9 | 3.6 | — | 4.2 | 3.9 | 4.2 | 2.8 | 0.8 |
| 3 | 1.8 | 0.6 | 2.8 | — | 2.9 | 3.1 | 3.5 | 2.4 | 0.6 |
| 2 | 1.4 | — | 1.9 | — | 1.7 | 2.4 | 2.7 | 2.0 | 0.4 |
| 1 | 1.0 | — | 1.0 | — | 0.5 | 1.7 | 1.9 | 1.6 | 0.2 |
| 0 | 0.6 | — | 0.1 | — | — | 1.0 | 1.1 | 1.2 | 0.0 |

## TABLE 2.

## THE INFLUENCE OF AGE ON MEAN WEIGHTED SCORES

(SMOOTHED)

| | N | 120 | 125 | 110 | 100 | 75 | 60 | 45 | 36 | 50 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AGE | | 20–24 | 25–29 | 30–34 | 35–39 | 40–44 | 45–49 | 50–54 | 55–59 | 60–69 | AVERAGE STANDARD DEVIATION |
| INFORMATION | | 10·4 | 10·2 | 10·1 | 10·0 | 9·8 | 9·7 | 9·5 | 9·4 | 9·3 | 3·29 |
| COMPREHENSION | | 10·0 | 10·2 | 10·0 | 9·8 | 9·6 | 9·4 | 9·3 | 9·1 | 8·9 | 3·01 |
| ARITHMETIC | | 9·8 | 9·6 | 9·4 | 9·2 | 8·9 | 8·7 | 8·5 | 8·3 | 8·0 | 3·19 |
| DIGIT SPAN | | 10·0 | 9·7 | 9·3 | 8·8 | 8·5 | 8·1 | 7·7 | 7·3 | 6·9 | 3·06 |
| SIMILARITIES | | 10·1 | 10·0 | 9·8 | 9·6 | 9·4 | 9·0 | 8·5 | 7·9 | 7·1 | 2·79 |
| PICTURE ARRANGEMENT | | 10·5 | 9·9 | 9·3 | 8·7 | 8·1 | 7·6 | 7·0 | 6·4 | 5·8 | 2·98 |
| PICTURE COMPLETION | | 10·2 | 9·8 | 9·4 | 9·1 | 8·7 | 8·4 | 8·0 | 7·6 | 7·3 | 3·14 |
| BLOCK DESIGN | | 10·8 | 10·3 | 9·6 | 9·0 | 8·5 | 7·9 | 7·3 | 6·7 | 6·1 | 3·10 |
| OBJECT ASSEMBLY | | 10·0 | 9·8 | 9·5 | 9·2 | 8·9 | 8·7 | 8·4 | 8·1 | 7·8 | 3·10 |
| DIGIT SYMBOL | | 10·9 | 10·3 | 9·5 | 8·8 | 8·0 | 7·2 | 6·5 | 5·8 | 5·0 | 2·98 |

1. The tables above and below involve decimals.
2. They overcome the grouping inherent in Wechsler's conversion which loses information.
3. A slight change in raw score is directly reflected in the decimal weighted score.
4. The supposedly equal steps of the raw score are not rendered unequal when converted.
5. Decimal weighted scores on one subtest are more nearly equivalent to those on another.
6. Their use may unmask trends within groups of individuals, and between intra-individual test-retest scores.

The "decimal weighted scores" on every subtest separately are then compared with the smoothed mean score for the particular age level (see Table 2)[2].

This comparison is aided by a ready-reckoner, the matrix of which is given in Table 3 for eight subtests only, Object Assembly and Digit Symbol being omitted for reasons of space, and Vocabulary since the Table of Contrived Means shown in Table 4 is not strictly comparable to the data given in Wechsler's Tables 39 and 40.

[2]The smoothing consisted in plotting mean scores at all age levels given in Wechsler's Tables 39 and 40, for every subtest separately, incorporating the figures given by Fox and Birren[7], and obtaining the lines of best fit.

TABLE 3. READY-RECKONER FOR COMPUTING RELATIONSHIP OF ANY WEIGHTED SCORE WITH MEAN AT THE AGE. (MEAN SCORES ARE IN *ITALICS*.)

| | 14-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-69 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **INFORMATION** | | | | | | | | | | | |
| | 17·5 | 18·7 | 18·5 | 18·4 | 18·3 | 18·1 | 18·0 | 17·8 | 17·7 | 17·6 | |
| | 16·1 | 17·0 | 16·8 | 16·7 | 16·6 | 16·4 | 16·3 | 16·1 | 16·0 | 15·9 | |
| | 14·7 | 15·4 | 15·2 | 15·1 | 15·0 | 14·8 | 14·7 | 14·5 | 14·4 | 14·3 | |
| | 13·3 | 13·7 | 13·5 | 13·4 | 13·3 | 13·1 | 13·0 | 12·8 | 12·7 | 12·6 | |
| | 11·9 | 12·1 | 11·9 | 11·8 | 11·7 | 11·5 | 11·4 | 11·2 | 11·1 | 11·0 | |
| M | 10·5 | 10·4 | 10·2 | 10·1 | 10·0 | 9·8 | 9·7 | 9·5 | 9·4 | 9·3 | M |
| | 9·1 | 8·8 | 8·6 | 8·5 | 8·4 | 8·2 | 8·1 | 7·9 | 7·8 | 7·7 | |
| | 7·7 | 7·1 | 6·9 | 6·8 | 6·7 | 6·5 | 6·4 | 6·2 | 6·1 | 6·0 | |
| | 6·3 | 5·5 | 5·3 | 5·2 | 5·1 | 4·9 | 4·8 | 4·6 | 4·5 | 4·4 | |
| | 4·9 | 3·8 | 3·6 | 3·5 | 3·4 | 3·2 | 3·1 | 2·9 | 2·8 | 2·7 | |
| | 3·5 | 2·2 | 2·0 | 1·9 | 1·8 | 1·6 | 1·5 | 1·3 | 1·2 | 1·1 | |
| | 2·1 | 0·5 | 0·3 | 0·2 | 0·1 | – | – | – | – | – | |
| **COMPREHENSION** | | | | | | | | | | | |
| | 17·1 | 17·6 | 17·8 | 17·6 | 17·4 | 17·2 | 17·0 | 16·9 | 16·7 | 16·5 | |
| | 15·6 | 16·0 | 16·2 | 16·0 | 15·8 | 15·6 | 15·4 | 15·3 | 15·1 | 14·9 | |
| | 14·2 | 14·5 | 14·7 | 14·5 | 14·3 | 14·1 | 13·9 | 13·8 | 13·6 | 13·4 | |
| | 12·7 | 13·0 | 13·2 | 13·0 | 12·8 | 12·6 | 12·4 | 12·3 | 12·1 | 11·9 | |
| | 11·3 | 11·5 | 11·7 | 11·5 | 11·3 | 11·1 | 10·9 | 10·8 | 10·6 | 10·4 | |
| M | 9·8 | 10·0 | 10·2 | 10·0 | 9·8 | 9·6 | 9·4 | 9·3 | 9·1 | 8·9 | M |
| | 8·3 | 8·5 | 8·7 | 8·5 | 8·3 | 8·1 | 7·9 | 7·8 | 7·6 | 7·4 | |
| | 6·9 | 7·0 | 7·2 | 7·0 | 6·8 | 6·6 | 6·4 | 6·3 | 6·1 | 5·9 | |
| | 5·4 | 5·5 | 5·7 | 5·5 | 5·3 | 5·1 | 4·9 | 4·8 | 4·6 | 4·4 | |
| | 4·0 | 4·0 | 4·2 | 4·0 | 3·8 | 3·6 | 3·4 | 3·3 | 3·1 | 2·9 | |
| | 2·5 | 2·5 | 2·7 | 2·5 | 2·3 | 2·1 | 1·9 | 1·8 | 1·6 | 1·4 | |
| | 1·0 | 0·9 | 1·1 | 0·9 | 0·7 | 0·5 | 0·3 | 0·2 | 0·0 | – | |
| **ARITHMETIC** | | | | | | | | | | | |
| | 15·6 | 17·8 | 17·6 | 17·4 | 17·2 | 16·9 | 16·7 | 16·5 | 16·3 | 16·0 | |
| | 14·3 | 16·2 | 16·0 | 15·8 | 15·6 | 15·3 | 15·1 | 14·9 | 14·7 | 14·4 | |
| | 12·9 | 14·6 | 14·4 | 14·2 | 14·0 | 13·7 | 13·5 | 13·3 | 13·1 | 12·8 | |
| | 11·6 | 13·0 | 12·8 | 12·6 | 12·4 | 12·1 | 11·9 | 11·7 | 11·5 | 11·2 | |
| | 10·2 | 11·4 | 11·2 | 11·0 | 10·8 | 10·5 | 10·3 | 10·1 | 9·9 | 9·6 | |
| M | 8·9 | 9·8 | 9·6 | 9·4 | 9·2 | 8·9 | 8·7 | 8·5 | 8·3 | 8·0 | M |
| | 7·6 | 8·2 | 8·0 | 7·8 | 7·6 | 7·3 | 7·1 | 6·9 | 6·7 | 6·4 | |
| | 6·2 | 6·6 | 6·4 | 6·2 | 6·0 | 5·7 | 5·5 | 5·3 | 5·1 | 4·8 | |
| | 4·9 | 5·0 | 4·8 | 4·6 | 4·4 | 4·1 | 3·9 | 3·7 | 3·5 | 3·2 | |
| | 3·5 | 3·4 | 3·2 | 3·0 | 2·8 | 2·5 | 2·3 | 2·1 | 1·9 | 1·6 | |
| | 2·2 | 1·8 | 1·6 | 1·4 | 1·2 | 0·9 | 0·7 | 0·5 | 0·3 | 0·0 | |
| | 0·9 | 0·2 | 0·0 | – | – | – | – | – | – | – | |
| **DIGIT SPAN** | | | | | | | | | | | |
| | 15·4 | 17·7 | 17·4 | 17·0 | 16·5 | 16·2 | 15·8 | 15·4 | 15·0 | 14·6 | |
| | 14·1 | 16·1 | 15·8 | 15·4 | 14·9 | 14·6 | 14·2 | 13·8 | 13·4 | 13·0 | |
| | 12·8 | 14·6 | 14·3 | 13·9 | 13·4 | 13·1 | 12·7 | 12·3 | 11·9 | 11·5 | |
| | 11·5 | 13·1 | 12·8 | 12·4 | 11·9 | 11·6 | 11·2 | 10·8 | 10·4 | 10·0 | |
| | 10·2 | 11·5 | 11·2 | 10·8 | 10·3 | 10·0 | 9·6 | 9·2 | 8·8 | 8·4 | |
| M | 8·9 | 10·0 | 9·7 | 9·3 | 8·8 | 8·5 | 8·1 | 7·7 | 7·3 | 6·9 | M |
| | 7·6 | 8·5 | 8·2 | 7·8 | 7·3 | 7·0 | 6·6 | 6·2 | 5·8 | 5·4 | |
| | 6·3 | 6·9 | 6·6 | 6·2 | 5·7 | 5·4 | 5·0 | 4·6 | 4·2 | 3·8 | |
| | 5·0 | 5·4 | 5·1 | 4·7 | 4·2 | 3·9 | 3·5 | 3·1 | 2·7 | 2·3 | |
| | 3·7 | 3·9 | 3·6 | 3·2 | 2·7 | 2·4 | 2·0 | 1·6 | 1·2 | 0·8 | |
| | 2·5 | 2·3 | 2·0 | 1·6 | 1·1 | 0·8 | 0·4 | 0·0 | – | – | |
| | 1·2 | 0·8 | 0·5 | 0·1 | – | – | – | – | – | – | |
| **SIMILARITIES** | | | | | | | | | | | |
| | 15·3 | 17·1 | 17·0 | 16·8 | 16·6 | 16·4 | 16·0 | 15·5 | 14·9 | 14·1 | |
| | 14·2 | 15·7 | 15·6 | 15·4 | 15·2 | 15·0 | 14·6 | 14·1 | 13·5 | 12·7 | |
| | 13·0 | 14·3 | 14·2 | 14·0 | 13·8 | 13·6 | 13·2 | 12·7 | 12·1 | 11·3 | |
| | 11·9 | 12·9 | 12·8 | 12·6 | 12·4 | 12·2 | 11·8 | 11·3 | 10·7 | 9·9 | |
| | 10·7 | 11·5 | 11·4 | 11·2 | 11·0 | 10·8 | 10·4 | 9·9 | 9·3 | 8·5 | |
| M | 9·6 | 10·1 | 10·0 | 9·8 | 9·6 | 9·4 | 9·0 | 8·5 | 7·9 | 7·1 | M |
| | 8·3 | 8·7 | 8·6 | 8·4 | 8·2 | 8·0 | 7·6 | 7·1 | 6·5 | 5·7 | |
| | 7·3 | 7·3 | 7·2 | 7·0 | 6·8 | 6·6 | 6·2 | 5·7 | 5·1 | 4·3 | |
| | 6·2 | 5·9 | 5·8 | 5·6 | 5·4 | 5·2 | 4·8 | 4·3 | 3·7 | 2·9 | |
| | 5·0 | 4·5 | 4·4 | 4·2 | 4·0 | 3·8 | 3·4 | 2·9 | 2·3 | 1·5 | |
| | 3·9 | 3·1 | 3·0 | 2·8 | 2·6 | 2·4 | 2·0 | 1·5 | 0·9 | 0·1 | |
| | 2·8 | 1·7 | 1·6 | 1·4 | 1·2 | 1·0 | 0·6 | 0·1 | – | – | |
| **PICTURE ARRANGEMENT** | | | | | | | | | | | |
| | 18·3 | 18·0 | 17·4 | 16·8 | 16·2 | 15·6 | 15·1 | 14·6 | 14·0 | 13·4 | |
| | 16·8 | 16·5 | 15·9 | 15·3 | 14·7 | 14·1 | 13·6 | 13·0 | 12·4 | 11·8 | |
| | 15·3 | 15·0 | 14·4 | 13·8 | 13·2 | 12·6 | 12·1 | 11·5 | 10·9 | 10·3 | |
| | 13·7 | 13·5 | 12·9 | 12·3 | 11·7 | 11·1 | 10·6 | 10·0 | 9·4 | 8·8 | |
| | 12·2 | 12·0 | 11·4 | 10·8 | 10·2 | 9·6 | 9·1 | 8·5 | 7·9 | 7·3 | |
| M | 10·7 | 10·5 | 9·9 | 9·3 | 8·7 | 8·1 | 7·6 | 7·0 | 6·4 | 5·8 | M |
| | 9·2 | 9·0 | 8·4 | 7·8 | 7·2 | 6·6 | 6·1 | 5·5 | 4·9 | 4·3 | |
| | 7·7 | 7·5 | 6·9 | 6·1 | 5·5 | 4·9 | 4·4 | 3·8 | 3·2 | 2·6 | |
| | 6·1 | 6·0 | 5·4 | 4·8 | 4·2 | 3·6 | 3·1 | 2·5 | 1·9 | 1·3 | |
| | 4·6 | 4·5 | 3·9 | 3·5 | 2·7 | 2·1 | 1·6 | 1·0 | 0·4 | – | |
| | 3·1 | 3·1 | 2·5 | 1·9 | 1·3 | 0·7 | 0·2 | – | – | – | |
| | 1·6 | 1·6 | 1·0 | 0·6 | 0·0 | – | – | – | – | – | |
| **PICTURE COMPLETION** | | | | | | | | | | | |
| | 16·7 | 18·1 | 17·7 | 17·3 | 17·0 | 16·6 | 16·3 | 15·9 | 15·5 | 15·2 | |
| | 15·3 | 16·5 | 16·1 | 15·7 | 15·4 | 15·0 | 14·7 | 14·3 | 13·9 | 13·6 | |
| | 13·9 | 14·9 | 14·5 | 14·1 | 13·8 | 13·4 | 13·1 | 12·7 | 12·3 | 12·0 | |
| | 12·5 | 13·3 | 12·9 | 12·5 | 12·2 | 11·8 | 11·5 | 11·1 | 10·7 | 10·4 | |
| | 11·1 | 11·8 | 11·4 | 11·0 | 10·7 | 10·3 | 10·0 | 9·6 | 9·2 | 8·9 | |
| M | 9·7 | 10·2 | 9·8 | 9·4 | 9·1 | 8·7 | 8·4 | 8·0 | 7·6 | 7·3 | |
| | 8·3 | 8·6 | 8·2 | 7·8 | 7·5 | 7·1 | 6·8 | 6·4 | 6·0 | 5·7 | |
| | 6·9 | 7·1 | 6·7 | 6·3 | 6·0 | 5·6 | 5·3 | 4·9 | 4·5 | 4·2 | |
| | 5·5 | 5·5 | 5·1 | 4·7 | 4·4 | 4·0 | 3·7 | 3·3 | 2·9 | 2·6 | |
| | 4·1 | 3·9 | 3·5 | 3·1 | 2·8 | 2·4 | 2·1 | 1·7 | 1·3 | 1·0 | |
| | 2·7 | 2·4 | 2·0 | 1·6 | 1·3 | 0·9 | 0·6 | 0·2 | – | – | |
| | 1·2 | 0·8 | 0·4 | 0·0 | – | – | – | – | – | – | |
| **BLOCK DESIGN** | | | | | | | | | | | |
| | 18·4 | 18·6 | 18·1 | 17·4 | 16·8 | 16·3 | 15·7 | 15·1 | 14·5 | 13·9 | |
| | 16·9 | 17·0 | 16·5 | 15·8 | 15·2 | 14·7 | 14·1 | 13·5 | 12·9 | 12·3 | |
| | 15·4 | 15·5 | 15·0 | 14·3 | 13·7 | 13·2 | 12·6 | 12·0 | 11·4 | 10·8 | |
| | 13·8 | 13·9 | 13·4 | 12·7 | 12·1 | 11·6 | 11·0 | 10·4 | 9·8 | 9·2 | |
| | 12·3 | 12·4 | 12·1 | 11·4 | 10·8 | 10·3 | 9·7 | 9·1 | 8·5 | 7·9 | |
| M | 10·8 | 10·8 | 10·3 | 9·6 | 9·0 | 8·5 | 7·9 | 7·3 | 6·7 | 6·1 | M |
| | 9·3 | 9·3 | 8·8 | 8·1 | 7·5 | 7·0 | 6·4 | 5·8 | 5·2 | 4·8 | |
| | 7·8 | 7·7 | 7·2 | 6·5 | 5·9 | 5·4 | 4·8 | 4·2 | 3·6 | 3·0 | |
| | 6·2 | 6·2 | 5·7 | 5·0 | 4·4 | 3·9 | 3·3 | 2·7 | 2·1 | 1·5 | |
| | 4·7 | 4·6 | 4·1 | 3·4 | 2·8 | 2·3 | 1·7 | 1·1 | 0·5 | – | |
| | 3·2 | 3·1 | 2·6 | 1·9 | 1·3 | 0·8 | 0·2 | – | – | – | |
| | 1·7 | 1·5 | 1·0 | 0·3 | – | – | – | – | – | – | |

The columns represent the same age groups from 14-69 as in Table 2. The mean weighted scores are shown in upright figures, and the rows above and below give the weighted scores corresponding to steps of half a standard deviation above and below the mean.[4]

A sliding cover may be prepared for the ready-reckoner. This has nine slots, one of which adjusts to the age group of the patient, the remainder uncovering the mean weighted scores on the various subtests. Horizontal lines on the cover are labelled "Mean", "1 S. D. above mean", and so on.

In use, the patient's weighted scores on all subtests are converted to *Sigma scores* or *Z scores*,[(2)] which can be corrected to the nearest quarter of a standard de-

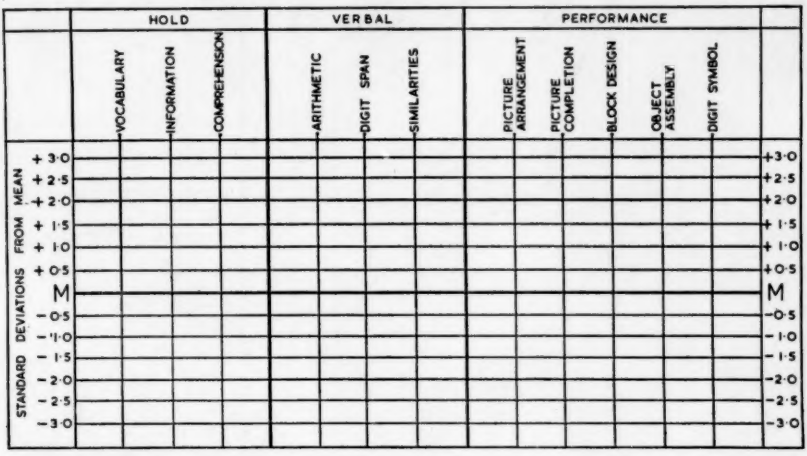
[4]The standard deviations (also shown in Table 2) for every subtest separately were averaged over all age groups, since despite the expectation that scatter would increase with age Wechsler[(30, Fig. 2 and Table 18 and Table 4 below)] they showed no trends, and much variability at different ages.

TABLE 4. APPROXIMATED VALUES FOR MEANS AND STANDARD DEVIATIONS ON VOCABULARY TEST (W-B I). (BY COURTESY OF DR. DAVID WECHSLER)

| Age | Mean | SD |
|---|---|---|
| 17–19 | 10.0 | 3.0 |
| 20–24 | 10.2 | 3.1 |
| 25–29 | 10.1 | 3.2 |
| 30–34 | 10.0 | 3.2 |
| 35–39 | 10.0 | 3.3 |
| 40–44 | 9.9 | 3.3 |
| 45–49 | 9.8 | 3.4 |
| 50–54 | 9.7 | 3.5 |
| 55–59 | 9.6 | 3.6 |

viation by interpolation. The scores are most conveniently presented to the clinician who referred the case as a scattergram, since it is easily read without the necessity of understanding the statistical implications, and strikes a good balance between over-sensitivity (which would emphasize normal intra-individual variability too much) and lack of sensitivity (which would mask genuine impairment). A suitable form is suggested in Fig. 1, the "hold" subtests being placed on the left. Compre-

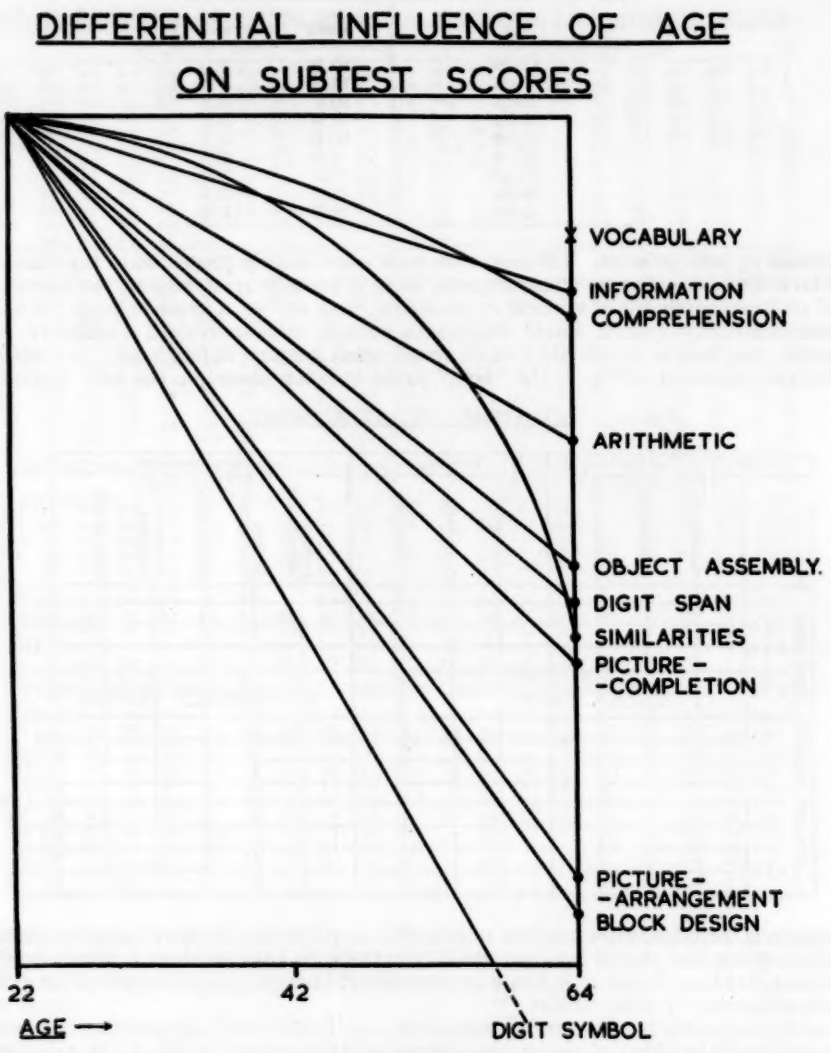
Fig. 1. SCATTERGRAM OF SIGMA SCORES




hension is included, although less reliable[25]. A difference of more than one sigma unit between the overall level on "hold" subtests, and the level on a "don't hold" subtest, has been found to indicate impairment of borderline significance of the abilities measured by that subtest.[10]

An incidental result from the compilation of Table 2 is the diagrammatic representation of the effect of age on the various subtests shown in Fig. 2. In this, the smoothed mean weighted scores at age 22 have been brought to a common origin, and their rates of decline plotted on the same scale up to age 64. A point representing Fox and Birren's *unsmoothed* data on Vocabulary for the age group 60-69 is given, although unreliable because there are no data for age 20-59 (except Wechsler's "contrived table" shown in Table 4). The rank order of subtests most affected by age is almost identical with those obtained by previous authors[6, 19, 26], especially since Fox and Birren[6] found that adjacent rankings were not significantly different at the 5% level.

FIG. 2.




SUMMARY

A review of the literature shows that deterioration indices based on weighted scores have been shown to be invalid, and that Vocabulary and Information scores provide the best baseline for scatter analysis. A technique is described in which a patient's scores are compared with the mean scores for his age. A difference of more than one sigma score (or Z score) between "hold" subtests, and any "don't hold" subtest is considered to indicate impairment on that subtest. As a corollary, the influence of age on mean subtest scores is illustrated.

## References


1. Anderson, A. L. Effect of laterality or localization of focal brain lesions on W-B subtests. *J. Clin. Psychol.*, 1951, 7, 149-53.
2. Barnett, I. The use of Z scores in equating W-B subtests. *J. clin. Psychol.*, 1950, 6, 184-188.
3. Botwinick, J. W-B split-half subtest reliabilities—differences in age and mental states. *J. consult. Psychol.*, 1952, 16, 225-228.
4. Corsini, R. and Fassett, K. Validity of Wechsler's deterioration index. *J. consult Psychol.*, 1952, 16, 462-468.
5. Derner, G., Aborn, M. and Canter, A. Reliability of the W-B subtests and scales. *J. consult. Psychol.*, 1950, 14, 172-179.
6. Fox, C. and Birren, J. E. Differential decline of subtest scores of the W-B Intelligence scale in 60-69 year old individuals. *J. genet. Psychol.*, 1950, 77, 313-17.
7. Fox, C. & Birren, J. E. Intellectual deterioration in the aged. *J. consult. Psychol.*, 1950, 14, 305-10.
8. Garfield, S. and Blek, L. Age, vocabulary level and mental impairment. *J. consult. Psychol.*, 1952, 16, 395-398.
9. Gilhooley, F. M. Relationship between variability and ability on the W-B scale. *J. consult. Psychol.*, 1950, 14, 46-48.
10. Gilhooley, F. M. W-B reliability and the validity of certain diagnostic signs of the neuroses. *J. consult. Psychol.*, 1950, 14, 82-87.
11. Glik, E. Comparison of recall and recognition types of measurement on verbal items. *J. clin. Psychol.*, 1951, 7, 157-162.
12. Gutman, B. Application of W. B. scale in diagnosis of organic brain disorders. *J. clin. Psychol.*, 1950, 6, 195-198.
13. Hamister, R. Test-retest reliability of W. B. I. for a neuro-psychiatric population. *J. consult. Psychol.*, 1949, 13, 39-43.
14. Hewson, L. The W. B. scale, and subtests as aids in neuro-psychiatric diagnosis. *J. nerv. ment. dis.*, 1949, 109, 158-183 and 246-266.
15. Hunt, W. L. Relative rates of decline of W. B. "hold" and "don't hold" tests. *J. consult. Psychol.*, 1949, 13, 440-443.
16. Klebanoff, S. G., Singer, J. L. and Wilensky, H. Psychological consequences of brain lesions and ablations. *Psychol. Bull.*, 1954, 51, 1-41.
17. McFie, J. and Piercey, M. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, 75, 292-311.
18. McCullough, M. W. B. changes following prefrontal lobectomy. *J. clin. Psychol.*, 1950, 6, 270-273.
19. Madonick, M. and Solomon, M. The W. B. scale in individuals past sixty. *Geriatics.*, 1947, 2, 34-40.
20. Marks, M. Criticism of the use of the W. B. scale as a diagnostic instrument. *J. gen. Psychol.*, 1953, 49, 143-152.
21. Markwell, E. and Wheeler, M. Changes in W. B. test performance following Prefrontal Lobectomy. *J. Consult. Psychol.*, 1953, 17, 229-231.
22. Merrill, R. and Heathers, L. Deviations on W. B. subtest scores from vocabulary level in University Counseling Center clients. *J. consult. Psychol.*, 1952, 16, 469-472.
23. Mettler, F. *Selective partial ablation of the frontal cortex.* New York: Hoeber, 1949.
24. Milner, B. Intellectual function of the temporal lobes. *Psychol. Bull.*, 1954, 51, 42-62.
25. Petrie, A. *Personality and the frontal lobes.* London: Routledge, Kegan & Paul, 1952.
26. Rabin, A. Psychometric trends in senility and psychosis of the senium. *J. gen. Psychol.*, 1944, 32, 149-162.
27. Rapaport, D. *Diagnostic Psychological testing.* Chicago: Year Book Co., 1945.
28. Reynell, W. A psychometric method of determining intellectual loss following head injury. *J. ment. sci.*, 1944, 90, 710-719.
29. Schalling, D. Qualitative changes in a Vocabulary test after Lobectomy and selective frontal operations. *Reports from Psychol. Lab. Stockholm. April, 1954.*
30. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins Co. 1944.
31. Wechsler, D. Personal Communication. 1954.
32. Wheeler, J. and Williams, W. The validity of the Hewson ratios. *J. consult. Psychol.*, 1951, 15, 163-165.
33. Wittenborn, J. and Holzberg, J. The W. B. and descriptive diagnosis. *J. consult. Psychol.*, 1951, 15, 325-329.

# THE EXPRESSIVE VALUE OF TOYS RECOMMENDED FOR NONDIRECTIVE PLAY THERAPY

DELL LEBO[1]

*Carter Memorial Hospital, Indianapolis*

## PROBLEM

The importance of toys has differed in the various kinds of play therapy. In some therapies toys have been the nucleus, in others of minimal importance[12]. Toys used in nondirective play therapy simply seem to have been derived from the creative toys of those with a Freudian theoretical orientation. The need for a better approach to the selection of toys in nondirective play therapy was noted as early as 1948[6]. A recent review of the literature of nondirective play therapy[10] did not indicate any experimental work on the question of toys. Actually, there has been no objective evidence that toys do any more than make the treatment hour enjoyable to the child.

A single experiment on the place of toys in nondirective play therapy[11], appearing after the review[10], suggested that there was no difference in the significance of children's verbalizations when they were playing or not playing with toys. That study can be criticized on several grounds. A more meaningful experiment would have included all of the children's statements, would not have relied upon undergraduate students for judges, and would have carefully differentiated between recommended toys, e.g., those appearing in the published lists of recognized therapists, and toys not on their lists. Because of the lack of objective information on the question of toys, a reexamination of the data is desirable.

The present paper proposes to test an assumption widely agreed upon in nondirective play therapy circles which is stated in the form of a specific hypothesis: The use of recommended toys will encourage the child to verbally express himself to a greater extent than will nonrecommended toys or no toys at all.

## METHOD

For purposes of the present experiment "using toys" is defined as manual manipulation of a toy, while "not using toys" means no manual manipulation of a toy. Visual orientation toward toys may have occurred in either category.

The following types of toys were designated as giving rise to spontaneous dramatic play, or creative free association by one or more therapists[1, 2, 15, 16]: Airplanes, a variety of animals, dolls, balloons, black board and chalk, boats, nursing bottles, bow and arrows, brooms, several types of cars, clay, cloth or rags, clothesline and clothespins, crayons, cups and saucers, hand puppets, doll beds, a doll house completely furnished and a doll family, guns, hatchets and knives of rubber, mallets, masks, paints of various kinds, paper dolls, peg board for pounding, pencils, doll bed equipment, sand, scissors, shovels, spoons, telephone, toy soldiers, water, and wooden sticks and blocks of different sizes. All these toys were employed in the present study as recommended toys.

With one exception the toys listed below were on none of the lists consulted. The single exception is the checker game. Axline mentions checker games in order to warn that like mechanical toys, they "are not the best type of material for expressive play"[1, p. 56]. The nonrecommended toys used in the present experiment included: Rubber balls, bubble blowing equipment, checker game, coffee pot, coloring books, comic books, cord or rope, film viewers with film strips, fireman's helmet, lady's hats, handcuffs, hoe, magazines, marbles, play money, paper pumpkin, rake, lady's shoes, stand-up-figures of Peter Rabbit, his family, and Farmer Brown, man's

[1]The writer wishes to thank: Dr. Glenn Terrell, Jr., Florida State University, for his helpful reading of this paper; Dr. Anders Sweetland, Florida State University, and James B. Morris, Milledgeville State Hospital, for serving as judges; Miss Janet Pierce and Mrs. Elaine Lebo, for assistance in establishing the factual material.

sweater, thumbtacks, washboard, and a whistle. The nonrecommended toys were all structured toys.

The above recommended and nonrecommended toys were available in the playroom for each play session.

*Subjects.* Twenty children of normal intelligence and social adjustment were used as subjects. An examination of the children's cumulative records and talks with their teachers determined their social adjustment. All subjects were within ± 1 sigma of 100 on the Stanford-Binet. Five ages were represented with two boys and two girls at each age level. Children were selected who did not vary more than four months from the age of four, six, eight, ten, or twelve years.

*Experimental Design.* Each subject was given three one hour individual nondirective play therapy sessions by the same therapist. As each child behaved in the playroom a verbatim style record was made of his statements and a descriptive style record was made of his actions. The statements, 4,692 for all subjects, were then categorized according to a method developed from nondirective work with children and described fully elsewhere [8, 13]. A list of the categories is given in Table 1. It has been indicated elsewhere [4, 8, 13] that the categories were reliably used by judges in differentiating the statements made by children in play therapy.

For the present experiment three experienced play therapists, including the present writer, used a five point rating scale to judge the categories on the basis of their indicativeness of the child's feelings or attitudes. A rating of one indicated that the judges considered a particular category to be very revealing of the child's feelings and attitudes. A rating of 5 meant that they considered the particular category to be very unrevealing of the child's feelings or attitudes. Of the 21 categories used in the present content analysis, two or more of the judges agreed upon 20. The only category on which all three judges disagreed in their rating was Q. This category was rated as four, three, and two by the judges.

Table 1 presents the mean ratings of the judges for each of the categories. In the case of 19 of the categories the mean value shown was also the modal value. Category C had a modal value of one and a mean value of two. The categories and the judge's ratings of the categories appear sufficiently reliable to warrent the use of chi-square. The problem of the independence of the measures should be treated [14]. It is true that since the children have each made more than one response per trial, inevitable linkages between their statements have resulted. However, the chi-square frequencies employed in the present experiment are categorizations of the statements, or ratings of the categories. Consequently, the separate events arrayed in the chi-square tables seem to be independent. The categorization of the third statement made by a child in a particular therapy session cannot appreciably influence the way a judge categorizes the fourth, or sixteenth, or any other statement.

Because there were more recommended than nonrecommended toys in the room, it was necessary to test the hypothesis by examining each of the categories by means of a chi-square analysis of per cents [3]. To equalize the number of statements made under the three conditions of play, it was necessary to convert the frequency of usage of each category to percentage figures. Fourteen of the categories did not have a sufficiently large theoretical frequency to validate the use of chi-square. Because of this the categories were also grouped according to the judges' ratings. That is to say, categories K, T, and U were grouped together because they were ranked as number one. The other categories were similarly grouped according to their collective ranks. This grouping enabled 19 of the categories to be subjected to a chi-square analysis.

## Results

The total percentages of use of each of the categories are shown in Table 1. In addition, the children's statements have been further divided into category percentages occurring when the children played with recommended toys, when they played

with nonrecommended toys, and when they spoke but were not playing with toys. These data are also available in Table 1.

TABLE 1. BRIEF DESCRIPTION OF CATEGORIES, JUDGE'S MEAN RATINGS, PER CENT OF TYPES OF STATEMENTS MADE IN THE PLAYROOM AND WHILE USING RECOMMENDED, NONRECOMMENDED, AND NO TOYS

| Categories | Judge's Mean Rating | % of Total Playroom Statements | Recommended Toys | Nonrecommended Toys | Not Playing With Toys |
|---|---|---|---|---|---|
| A. Curiosity | 2 | 4.96 | 3.26 | 2.75 | 11.05 |
| B. Description | 3 | 24.85 | 29.43 | 31.84 | 7.83 |
| C. Aggression | 2 | 2.66 | 3.30 | .52 | 2.36 |
| D. Story Units | 2 | 3.91 | 5.70 | .86 | .86 |
| E. Decisions | 2 | 2.98 | 3.64 | 3.61 | .75 |
| F. Doubt | 2 | 3.64 | 3.53 | 3.27 | 4.18 |
| G. Exploring Limits | 2 | 3.18 | 3.64 | 3.79 | 1.50 |
| H. Shift Responsibility | 2 | 1.78 | 2.17 | 1.03 | 1.18 |
| I. Interest | 2 | 2.25 | .85 | 2.07 | 6.22 |
| J. Attempting Relations | 3 | 8.24 | 8.80 | 7.92 | 6.87 |
| K. Negative about self | 1 | .78 | .81 | 1.20 | .43 |
| L. Positive about self | 2 | .39 | .43 | .69 | .11 |
| M. Negative about Others | 2 | 6.67 | 7.33 | 6.88 | 4.72 |
| N. Positive about Others | 2 | 3.89 | 3.45 | 3.79 | 5.15 |
| O. Information | 3 | 9.46 | 5.74 | 7.23 | 21.14 |
| P. Asking for Information | 4 | 5.72 | 6.01 | 6.37 | 4.51 |
| Q. Time Questions | 3 | 1.89 | .74 | 1.03 | 5.58 |
| R. Exclamations | 3 | 4.37 | 3.92 | 6.71 | 4.18 |
| S. Unclassifiable | 4 | 8.33 | 7.25 | 8.43 | 11.27 |
| T. Insight | 1 | —— | —— | —— | —— |
| U. Ambivalence | 1 | .002 | —— | —— | .11 |
| Total | | 99.95 | 100.00 | 99.99 | 100.00 |

The W statistic [7] was applied to the rank order of each category when recommended, unrecommended, or no toys were being used. A coefficient of concordance of .79 resulted. This figure was found to be significant at better than the one per cent level of confidence by Kendall's method [7]. The W statistic indicated a relatively high similarity in the rank order of use of each category for the present subjects. There seems to be a similarity in the kinds of statements made by children when they are playing with recommended or unrecommended toys or not playing with toys.

A chi-square analysis of each of the categories with a theoretical frequency of five or above was next performed. The W statistic had suggested that there was no difference in the percentages of use of the various categories. Consequently the theoretical frequencies could be drawn from the total percentage of use of each category from Table 1.

The seven categories with a sufficiently high theoretical frequency to be individually examined by chi-square are shown in Table 2, together with results of the tests.

TABLE 2. THE CHI-SQUARE VALUES AND SIGNIFICANCE LEVELS OF STATEMENT CATEGORIES

| Categories | Chi-Square | P |
|---|---|---|
| A. Curiosity | 9.04 | Better than .02 |
| B. Description | 136.31 | Better than .01 |
| J. Attempting Relations | .76 | Nonsignificant |
| M. Negative about Others | 1.47 | Nonsignificant |
| O. Information | 30.27 | Better than .01 |
| P. Asking for Information | .66 | Nonsignificant |
| S. Unclassifiable | 2.78 | Nonsignificant |

An examination of Tables 1 and 2 reveals that statements falling into categories A and O were more likely to be made when the children were not playing with toys then when playing with toys. Category A relates to curiosity about the situation and things present in it. Category O deals with straight information and stories about the family, school, pets, teachers, self, etc. Tables 1 and 2 reveal further that categories J, M, P, and S occurred with approximately equal frequency when recommended, unrecommended, or no toys were used by the children. These categories pertain to attempts to establish a relationship with counselor, negative statements about the family, school, things made or present in the playroom, the situation, activities, etc., asking for information, and unclassifiable statements, respectively.

Of all the categories in Table 1 and 2 with a sufficiently large theoretical frequency for an individual chi-square analysis, category B is the only one that was made more frequently while the children were playing with toys. Category B concerns statements of simple description, information, and comments about play and playroom. Whether the children were playing with recommended or nonrecommended toys did not seem to affect the use of this category.

In order to analyze more of the categories by the chi-square method they were grouped according to their expressive ratings. As can be seen from an examination of Table 1, three categories were rated as one, or very revealing. Their combined expected frequencies were not sufficient to warrant the employment of chi-square. Eleven categories were rated as two, or revealing. The chi-square of these categories was 22.01. This figure was significant at better than the one per cent level of confidence. The analysis indicated that nonrecommended toys were the least productive of such category use. However, the use of no toys was as effective as were recommended toys among the categories rated as revealing of children's feelings.

Five categories rated as three, or neither revealing nor unrevealing, yielded a chi-square of 18.37. This was significant at better than the one per cent level. Nonrecommended toys were found to be most productive of statements falling into these categories. Two categories rated as 4, or unrevealing, produced a chi-square of 1.71. This figure was nonsignificant. It suggested that nonrecommended toys, recommended toys and not playing with toys all resulted in an approximately equivalent usage of categories judged as unrevealing.

The hypothesis that the use of recommended toys will encourage the child to verbally express himself to a greater extent than will nonrecommended toys or no toys at all has not been supported with respect to any of the categories of this study.

## Discussion

Examination of the results reveals that the children's use of recommended toys was not invariably accompanied by the production of a significantly increased number of statements. This finding suggests that the lists of toys published by nondirectivists[1, 15] may be in error in stressing the type of toys used to encourage expression in psychoanalytically oriented play relationships. It is possible that the nondirective relationship established between the therapist and the child cannot employ expressive or creative toys exclusively. The nondirective relationship may require other types of toys. At our present state of knowledge concerning the question of toys in nondirective play therapy, Ruth Hartley's comment seems especially appropriate, ". . . . For each child the materials he uses have unique values dependent upon associations with his past and on his ability to project meanings and use symbols" [6, p. iii].

Client-centered therapy with adults gradually evolved from an eclectic usage of previous methods of psychotherapy and finally emerged as an independent school of thought[9]. In the eyes of the present writer nondirective play therapy must go through the same process. It seems anachronistic to have a modern client-centered philosophy underlying nondirective play therapy and to rely on toys apparently handed down from Freudian play therapy. Moustakas[15] may have been building wisely when he suggested that a variety of both structured and unstructured play

materials be employed in nondirective play therapy. However, he may not have gone far enough. He simply seems to have added a few toys such as balloons, scissors, shovels, etc. to Axline's[1] predominantly diagnostically centered list.

## Summary

Using 4,092 statements made in 60 individual nondirective play therapy sessions by 20 normal children aged 4 to 12 years the present experiment found that the use of toys suggested by nondirective play therapists did not seem to encourage children to verbally express themselves to a greater extent than did nonrecommended toys or not playing with toys. It is suggested that the nondirective play relationship might not require the creative or dramatic toys of the Freudian diagnostic play relationship.

## References

1. Axline, Virginia M. *Play therapy.* Boston: Houghton Mifflin, 1947.
2. Arthur, Helen. A comparison of the techniques employed in psychotherapy and psychoanalysis of children. *Amer. J. Orthopsychiat.*, 1952, *22*, 484-498.
3. Edwards, A. L. *Statistical analysis.* New York: Rinehart, 1946.
4. Finke, Helene. Changes in the expression of emotionalized attitudes in six cases of play therapy. Unpublished Master's thesis, Univer. of Chicago, 1947.
5. Harms, E. Play diagnosis: Preliminary considerations for a sound approach. *Nerv. Child*, 1948, *1*, 233-246.
6. Hartley, Ruth E. *Growing through play: experiences of Teddy and Bud.* New York: Columbia University Press, 1952.
7. Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.
8. Lebo, D. The relationship of response categories in play therapy to chronological age. *J. Child Psychiat.*, 1952, *2*, 330-336.
9. Lebo, D. The development of client-centered therapy in the writings of Carl Rogers. *Amer. J. Psychiat.*, 1953, *110*, 104-109.
10. Lebo, D. The present status of research on nondirective play therapy. *J. consult. Psychol.*, 1953, *17*, 177-183.
11. Lebo, D. The contribution of toys to the nondirective play therapy process. *J. Child Psychiat.*, in press.
12. Lebo, D. The development of play as a form of therapy: from Rousseau to Rogers. *Amer. J. Psychiat.*, in press.
13. Lebo, D. Quantification of the nondirective play therapy process. *J. genet. Psychol.*, in press.
14. Lewis, D. and Burke, C. J. The use and misuse of the chi-square test. *Psychol. Bull.*, 1949. *46*, 433-489.
15. Moustakas, C. E. *Children in play therapy.* New York: McGraw-Hill, 1953.
16. Watson, R. I. *The clinical method in psychology.* New York: Harper, 1951.

# AN ITEM ANALYSIS OF THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

FREDERICK O. CARLETON AND CHALMERS L. STACEY

*Syracuse University*

## Problem

The Wechsler Intelligence Scale for children has received increasing acceptance for evaluating the general intelligence level of children. Psychologists in state schools for the mentally defective, in particular, have looked upon the scale as showing promise as a useful instrument for evaluation and diagnosis. However, relatively little research data is currently available which is directly relevant to the performance of subnormal groups, especially in regard to differential item responses. Information relating to the nature of the individual item statistics obtained when the WISC has been applied to a representative sample of mental defectives would seem to be of value for a number of reasons. One is that psychologists, in addition to considering total performance, often make clinical inferences based upon particular item responses. To make meaningful inferences, we should know something of the typical response to such items for comparable subjects. In addition, psychologists are interested in a scale which cannot only differentiate between normal and subnormal groups but can also reveal individual differences within the special group. For this it is necessary to know something of the discriminating power of the individual items as it is recognized that total test performance is a function of the aggregate performance on the individual items comprising the scale. The order of difficulty of items when the scale is applied to the special group is also of interest as total test scores may be materially affected by the order of presentation of items in the testing situaton.

Item analysis studies for various normal and abnormal groups have been reported by Jastak[1] and Rapaport[4] in the case of the Wechsler-Bellevue Scale. These analyses of the W-B, which is comparable in item content to the WISC, have revealed item statistics which to some degree are at variance with those found by Wechsler. From his results, Jastak suggested a change in the order of test administration and revision of the record blank of the W-B scale. Such differences in item responses from that of the standardization group could also reasonably be expected from analysis of the WISC, especially when applied to subjects from a population which differs substantially from the standardization sample.

Data concerning the standardization of the WISC have been reported by Seashore[5] and Wechsler[6]. They report that the standardization sample included 55 institutionalized subjects classified as "feebleminded", a group which comprised 2.5 per cent of the total sample. These feebleminded subjects were rated as having IQs under 70 and not below 50. In addition, the standardization group included 399 school children whose full scale IQs were less than 90. Thus of the 2200 children in the standardization sample, 454 or 21 per cent obtained IQ's less than 90. We may, therefore, question the relevance of the WISC scale as presently constituted as a device for evaluating the intellectual level of children from state school populations.

The purpose of the present paper is to report obtained statistics relating to item response to the WISC for a sample of 366 subjects tested at the Syracuse State School who can be classified as defective, borderline and dull normal. This sample is considered to be fairly representative of subjects typically encountered in state institutions to whom the WISC would likely be administered.

## Sample and Procedure

During the past four years, the WISC has been routinely administered as part of a psychological test battery to subjects at the Syracuse State School. Subjects referred by social agencies as possible mental defectives were evaluated by the state

school staff during a sixty day observation period. Some of the subjects were subsequently recommended for institutional placement while others were not. The sample, described in Table 1, was selected from the total number of subjects tested at the school. Selection was made on the basis of completeness of test data as well as a diagnosis of freedom from organic involvement.

TABLE 1. MEAN, STANDARD DEVIATION AND RANGE IN WISC IQ, WISC WEIGHTED SCORE AND CHRONOLOGICAL AGE FOR 366 SUBJECTS

| Factors | Mean | S.D. | Range |
|---|---|---|---|
| WISC full scale IQ | 67.36 | 10.58 | 42–91 |
| WISC total weighted score | 55.26 | 10.33 | 22–88 |
| Chronological Age (yrs.) | 12.30 | 1.82 | 7.50–15.92 |

The sample consisted of 194 boys and 172 girls between the ages of seven years six months and fifteen years eleven months with a mean full scale IQ of 67.36. Although differential sex response to the various items may to some extent bias the results to be obtained, it was decided to combine boys and girls into one sample. This was done to increase the sample size and in so doing it was hoped to improve the stability of the derived statistics. In future studies, when a greater number of subjects is available, it is expected to make separate analyses for the two sexes. The desirability of such separate analyses is suggested by Jastak's data which revealed substantial differences in item responses for the two sexes.

Item responses, for each subject, were transferred from the WISC record blank to a standard two alternative IBM answer sheet.[1] The item responses were dichotomized according to a criterion of "right" vs. "wrong" in accordance with scoring instructions specified by the WISC manual [6]. Partial or qualitative scores, however, were not separately recorded. Items were considered to have been correctly answered if any scoring credit could be allowed and were accordingly recorded on the IBM answer sheet on an "all or none" basis. While this procedure may have resulted in reducing individual differences in item responses, it was thought desirable to increase the ordinal consistency of the test items as well as to simplify the analysis. Frequency counts for the individual items were obtained by use of the IBM graphic item counter.

Item validities were estimated by computing the biserial correlation coefficient between dichotomized item responses and the total weighted score. Computation of item validities was facilitated by making use of the Kolbe and Edgerton [3] table for estimating biserial correlation coefficients. The total weighted score was normalized by dividing the distribution into equal-frequency eighths and assigning the standard score equivalent of the midpoint of each eighth in a normal distribution to each criterion score within that eighth. Thus, all cases falling within a given eighth of the obtained distribution of total weighted scores received the same standard score equivalent. This method of analysis was chosen as it takes into account the total distribution of scores and consequently is expected to result in more stable item statistics than would be the case if scores selected from the tails of the distribution were used. The choice of a method which would maximize the reliability of the derived statistics seemed to be particularly desirable considering the relatively small size of the sample which had been chosen from a homogeneous population.

## RESULTS

Results of the analysis are presented in Table 2 which has been tabulated in three columns: (1) the frequency $f$ of subjects in the sample who answered the item correctly, (2) the proportion $p$ of the subjects who answered the item correctly, and

[1]The authors express their appreciation to Miss Mary Lowenberg who transferred item response from the record blanks to IBM answer sheets.

(3) the biserial correlation coefficient $r_{bis}$ between each item and the total weighted score. Items are presented in Table 2 in the same sequence as they appear in the WISC record blank. However, only those items are listed for which at least one person in the sample responded correctly. Thus, for example, only the first eighteen items of the information subtest appear in the table, as no subject responded correctly to the remaining twelve.

For each subtest, Kendall's tau (2) has been computed indicating the rank order correlation between the order of presentation according to Wechsler's manual and the order of difficulty obtained from the present sample. These rank order correlation coefficients appear in Table 2 for each subtest.

TABLE 2. FREQUENCY OF CORRECT RESPONSE (f). INDEX OF DIFFICULTY (p), AND INDEX OF DISCRIMINATION ($r_{bis}$) FOR ITEMS OF THE WISC

| Item | f | p | $r_{bis}$ |
|---|---|---|---|
| | Information | | |
| 1 | 361 | 97 | .21* |
| 2 | 359 | 97 | .14 |
| 3 | 348 | 94 | .27*** |
| 4 | 357 | 96 | .14 |
| 5 | 288 | 78 | .34*** |
| 6 | 280 | 76 | .30*** |
| 7 | 297 | 80 | .46*** |
| 8 | 245 | 66 | .42*** |
| 9 | 160 | 43 | .26*** |
| 10 | 195 | 53 | .41*** |
| 11 | 135 | 36 | .41*** |
| 12 | 103 | 28 | .11* |
| 13 | 44 | 12 | .32*** |
| 14 | 17 | 5 | .29*** |
| 15 | 21 | 6 | .23** |
| 16 | 4 | 1 | .29* |
| 17 | 6 | 2 | −.06 |
| 18 | 5 | 1 | .35** |
| Kendall's Tau .914 | | | |
| | Comprehension | | |
| 1 | 352 | 95 | .14* |
| 2 | 310 | 84 | .26*** |
| 3 | 255 | 69 | .23*** |
| 4 | 244 | 66 | .33*** |
| 5 | 185 | 50 | .49*** |
| 6 | 238 | 64 | .35*** |
| 7 | 92 | 25 | .33*** |
| 8 | 26 | 7 | .44*** |
| 9 | 30 | 8 | .50*** |
| 10 | 9 | 2 | .46*** |
| 11 | 2 | 1 | .28* |
| 12 | 1 | 0 | .00 |
| 13 | 0 | 0 | .00 |
| 14 | 1 | 0 | .00 |
| Kendall's Tau .939 | | | |

| Item | f | p | $r_{bis}$ |
|---|---|---|---|
| | Arithmetic | | |
| 1 | 355 | 96 | .26** |
| 2 | 352 | 95 | .33*** |
| 3 | 335 | 90 | .38*** |
| 4 | 312 | 84 | .31*** |
| 5 | 286 | 77 | .45*** |
| 6 | 185 | 50 | .51*** |
| 7 | 133 | 36 | .36*** |
| 8 | 98 | 26 | .46*** |
| 9 | 59 | 16 | .40*** |
| 10 | 2 | 1 | .28* |
| 11 | 7 | 2 | .40*** |
| 12 | 8 | 2 | .41*** |
| Kendall's Tau. .955 | | | |
| | Similarities | | |
| 1 | 330 | 89 | .16** |
| 2 | 312 | 84 | .28*** |
| 3 | 353 | 95 | −.13 |
| 4 | 155 | 42 | .28*** |
| 5 | 218 | 59 | .44*** |
| 6 | 193 | 52 | .46*** |
| 7 | 173 | 47 | .37*** |
| 8 | 213 | 58 | .45*** |
| 9 | 138 | 37 | .44*** |
| 10 | 21 | 6 | .40*** |
| 11 | 9 | 2 | .46*** |
| 12 | 1 | 0 | .00 |
| Kendall's Tau .879 | | | |
| | Digit Span (Forward) | | |
| 3 | 363 | 98 | .07 |
| 4 | 334 | 90 | .29*** |
| 5 | 194 | 52 | .35*** |
| 6 | 74 | 20 | .43*** |
| 7 | 16 | 4 | .14 |
| Kendall's Tau 1.00 | | | |

| Item | f | p | $r_{bis}$ |
|---|---|---|---|
| | Digit Span (backward) | | |
| 2 | 321 | 87 | .48*** |
| 3 | 247 | 67 | .47*** |
| 4 | 75 | 20 | .34*** |
| 5 | 11 | 3 | .57*** |
| Kendall's Tau 1.00 | | | |
| | Mazes | | |
| A | 342 | 92 | .32*** |
| B | 325 | 88 | .40*** |
| C | 317 | 86 | .41*** |
| 1 | 278 | 75 | .53*** |
| 2 | 261 | 70 | .51*** |
| 3 | 189 | 51 | .53*** |
| 4 | 195 | 53 | .51*** |
| 5 | 120 | 32 | .37*** |
| Kendall's Tau .929 | | | |
| | Vocabulary | | |
| 1 | 355 | 96 | .23** |
| 2 | 351 | 95 | .25** |
| 3 | 353 | 95 | .06 |
| 4 | 256 | 69 | .20*** |
| 5 | 347 | 94 | .16* |
| 6 | 300 | 81 | .24*** |
| 7 | 318 | 86 | .11* |
| 8 | 298 | 80 | .26*** |
| 9 | 218 | 59 | .34*** |
| 10 | 100 | 27 | 30*** |
| 11 | 210 | 57 | .32*** |
| 12 | 181 | 49 | .34*** |
| 13 | 196 | 53 | .39*** |
| 14 | 183 | 49 | .28*** |
| 15 | 118 | 32 | .39*** |
| 16 | 80 | 22 | .33*** |
| 17 | 69 | 19 | .41*** |
| 18 | 141 | 38 | .41*** |
| 19 | 12 | 3 | .20* |
| 20 | 24 | 6 | .35*** |
| 21 | 14 | 4 | .35*** |
| 22 | 2 | 1 | .34** |
| 23 | 1 | 0 | .00 |
| 24 | 3 | 1 | .10 |
| 25 | 41 | 1 | .63*** |
| Kendall's Tau .875 | | | |

***Significant at .01 level of confidence
** " " .05 " " "
* " " .10 " " "

TABLE 2 (continued).

| Item | f | p | rbis |
|---|---|---|---|
| **Picture Completion** | | | |
| 1 | 365 | 99 | .08 |
| 2 | 358 | 97 | .11 |
| 3 | 363 | 98 | .08 |
| 4 | 347 | 94 | .31*** |
| 5 | 326 | 88 | .28*** |
| 6 | 250 | 68 | .32*** |
| 7 | 310 | 84 | .38*** |
| 8 | 137 | 37 | .33*** |
| 9 | 167 | 45 | .37*** |
| 10 | 132 | 36 | .39*** |
| 11 | 152 | 41 | .43*** |
| 12 | 152 | 41 | .43*** |
| 13 | 39 | 11 | 35*** |
| 14 | 89 | 24 | 41*** |
| 15 | 20 | 5 | .28*** |
| 16 | 19 | 5 | .25** |
| 17 | 40 | 11 | .42*** |
| 18 | 4 | 1 | .12 |
| 19 | 28 | 8 | .33*** |
| 20 | 3 | 1 | .63*** |
| Kendall's Tau .851 | | | |

| Item | f | p | rbis |
|---|---|---|---|
| **Picture Arrangement** | | | |
| A | 364 | 98 | −.04 |
| B | 360 | 97 | .21* |
| C | 322 | 87 | .23*** |
| D | 205 | 55 | .45*** |
| 1 | 182 | 49 | .53*** |
| 2 | 166 | 45 | .55*** |
| 3 | 195 | 53 | .62*** |
| 4 | 108 | 29 | .58*** |
| 5 | 39 | 11 | .48*** |
| 6 | 37 | 10 | .42*** |
| 7 | 19 | 5 | .37*** |
| Kendall's Tau. 964 | | | |

| Item | f | p | rbis |
|---|---|---|---|
| **Block Design** | | | |
| A | 350 | 95 | .21** |
| B | 348 | 94 | .34*** |
| C | 326 | 88 | .43*** |
| 1 | 163 | 44 | .59*** |
| 2 | 54 | 15 | .55*** |
| 3 | 65 | 18 | .56*** |
| 4 | 79 | 21 | .67*** |
| 5 | 23 | 6 | .48*** |
| 6 | 9 | 2 | .46*** |
| 7 | 5 | 1 | .39** |
| Kendall's Tau .933 | | | |
| **Object Assembly** | | | |
| M | 361 | 97 | .08 |
| H | 343 | 93 | .36*** |
| F | 333 | 90 | .27*** |
| A | 340 | 92 | .23*** |
| Kendall's Tau 1.00 | | | |

*Order of Difficulty.* Rank order correlations (*tau*) computed for each of the various subtests range from .851 to 1.00 indicating that the obtained orders of difficulty tend to approximate rather closely the orders specified by the manual. Inspection of the *p* column of Table 2 would seem to substantiate the impression that relatively few items are misplaced in respect to order of presentation and that such misplacements as do occur do not seem to be of sufficient extent to affect materially the subtest total score. Thus, for this population the rearrangement of items from the usual WISC sequence to one made to conform with the order of difficulty found in this study would not seem warranted.

However, while the obtained orders of difficulty are not greatly different from the usual order of presentation, the present indices do differ markedly in magnitude from what would have been expected from a more nearly normal group. For each subtest there is a relatively abrupt shift from items which appear to be quite easy for the group to ones which are quite difficult. Relatively few items approach the ideal 50-50 split which would contribute most to total test variance. Therefore, it cannot be expected that the subtests would be particularly sensitive to revealing individual differences among subjects chosen from a comparable population. It should be pointed out, however, that the average index of difficulty for the test items is 47 when considering the items listed in Table 2 and 37 when considering all of the items listed in the record blank.

Table 3 presents the distribution of item difficulties for the 183 scorable items. From this it can be seen that 41 items or 22 percent of the total are completely nonfunctional as no subject in the group was successful in responding to them. The table also shows that the obtained distribution of item difficulties is far from normal and that a large number of the items contribute relatively little to the total test variance. Only 40 items are within the difficulty range of 30 to 70 which is considered by many test constructors to be most desirable. On the other hand, 57 per cent of the items are in the extreme deciles (0-10 and 90-100).

*Item Validities.* Item validity, a measure of item discriminating power, may be inferred from the degree of relationship between item responses and some criterion of performance. Biserial correlation coefficients indicating the degree of relationship between each item and the total weighted score are presented in Table 2 in the

column headed $r_{bis}$. Item validity coefficients which differ significantly from zero at the .01, .05, and .10 levels of confidence for the various values of $p$ and $q$ have been indicated in Table 2.

TABLE 3. DISTRIBUTION OF ITEM DIFFICULTIES

| Index of difficulty | Frequency |
|---|---|
| 90–99 | 28 |
| 80–89 | 15 |
| 70–79 | 5 |
| 60–69 | 7 |
| 50–59 | 14 |
| 40–49 | 11 |
| 30–39 | 8 |
| 20–29 | 10 |
| 10–19 | 9 |
| 0–9 | 76 (41)* |
| Total | 183 |

*Number in ( ) indicates number of items for which frequency of correct response was less than 1.

Of the 142 correlations listed in Table 2, 124 are significantly different from zero at the .10 level of confidence and of these 104 are significant at the .01 level. That such a high proportion of the items should show significant validity is a tribute to the extreme care taken in the original construction and standardization of the scale. A number of factors could have been expected to have reduced appreciably the obtained relationships. Among these is the difference in intellectual level and homogeneity of the present sample as well as the combining of test responses of boys and girls into one distribution. However, despite these known biases and other unknown biases which might tend to reduce the item-total test correlations, the obtained correlations are sufficiently strong to indicate that most of the items are operating in a discriminating manner. Therefore, even with the present homogeneous group of subjects most of the items appear to operate as valid predictors of general intelligence. This interpretation, of course, is dependent upon the assumption that the total test score is a valid measure of the intelligence level of the subjects.

## SUMMARY

Indices of item difficulty and item discrimination have been presented for the items comprising the Wechsler Intelligence Scale for children as obtained from a group of 366 suspected mental defectives. While item difficulties are roughly in the same rank orders as those obtained with the standardization sample, they are not distributed normally and there appear to be relatively few items of a moderate level of difficulty. For the present sample, it is likely that a large proportion of the items do not contribute significantly to total test variance. However, evidence relating to the internal consistency of the test suggests that most of the items operate as significant discriminators. The efficiency of the items as discriminators, however, is greatly reduced as many items apply to only small proportions of the sample. For more effective evaluation of mental defectives, it might be desirable to eliminate many of the present items which appear to be either too easy or too difficult and in their place substitute items of comparable validity whose difficulty indices are between 30 and 70. The development of a scale comprised of such items might be the answer to the desire of state school psychologists for a shorter test which will reveal individual differences among mental defectives. Until the time such a scale is developed, it would seem fruitful to continue application of the WISC as presently constituted. Changes in the order of test administration or revision of the record blank would not seem warranted on the basis of the present evidence.

### Bibliography

1. Jastak, Joseph. An item analysis of the Wechsler-Bellevue tests. *J. consult. psychol.* 1950, 14, 88-94.
2. Kendall, M. G. *Rank Correlation Methods.* London: Charles Griffen, 1948.
3. Kolbe, L. E. and Edgerton, H. A. A table for computing bi-serial *r*. *J. exp. educ.*, 1936, 4, 235-251.
4. Rapaport, D. et al. *Diagnostic psychological testing*, Vol. 1. Chicago: Year Book Publishers, 1945.
5. Seashore, H., Wesman, A. and Doppolt, J. The standardization of the Wechsler Intelligence Scale for Children. *J. consult. psychol.* 1950, 14, 99-110.
6. Wechsler, D. *Wechsler Intelligence Scale for Children* (manual). New York: Psychological Corporation, 1949, pp. 113.


## USE OF THE WISC IN PREDICTING ORGANICITY*


HARRY S. BECK
*Illinois Department of Public Instruction*

ROBERT L. LAM, M. D.
*Washington University Medical School*


### Problem

Since the etiological aspects of mental retardation in children are of prime importance in setting up an adequate educational program, the value of tests which will quantitatively differentiate organics from non-organics is readily apparent. The present study attempts to determine whether the WISC is useful in this respect. Two specific questions were considered:

1. Would the IQ's obtained on the various scales of the WISC significantly differentiate between organics and non-organics?
2. Would the WISC yield a pattern of subtest scores characteristic of organics only?

Wechsler[14] has indicated that the W-B Scale shows a patterning of subtest scores which is suggestive of organicity, but nothing has been reported on the WISC.

### Procedure

The subjects for this study were examined with psychological tests by one of us (H. S. B.) and consisted of 104 children between the ages of 6-0 and 15-11 years. All of these children were given the WISC, the Bender-Gestalt, and the H-T-P. In addition some of the children were given Rorschach tests, Yacorzynski's Test for Frontal Lobe Involvement, and the Color-Form Sorting Test. A rather detailed developmental and social history was also obtained for each child. The psychologist's decision to refer the child for neurological evaluation was based upon the over-all analysis of the test results and the history. The subjects were referred to the psychologist for examination by the public schools in Southern Illinois in order to determine their eligibility for placement in special classes for the educable mentally handicapped. They were all retarded in school, and many presented behavior problems. The majority of the subjects were from low socio-economic backgrounds, with many of them coming from broken homes.

The subjects have been placed in one of three groups according to their classification. One group consisted of 27 subjects with known neurological involvement, designated as the "organic group." The term "organic" is defined as: any involvement of the central nervous system resulting from structural changes caused by disease, injury, or malformation occurring before, during, or after birth. This group

*The authors wish to acknowledge their appreciation of the suggestions and statistical aid afforded by Doctors Marshall S. Hiskey, Chairman, and Eugene D. Fitzpatrick, Assistant Professor, Department of Guidance and Special Education, Southern Illinois University.

was examined neurologically and classified by one of us (R. L. L.). Included in this group were cases of cerebral dysrhythmia with and without clinical epilepsy, encephalitis, congenital choreo-athetosis, cerebral palsy, cerebral dysplasia or agenesis, and subdural hematoma. It may be noted at this point that Wechsler[14] did not limit his organics to any one type in evolving his pattern.

The second group consisted of 48 cases who had been referred for neurological examination because organicity was inferred by psychological studies. This was our "suspected group."

The third group was composed of 29 cases with no evidence of organicity from psychological evaluation and developmental history. For biometric completeness this group should have been examined neurologically, but due to financial and other reasons this was impossible at the time. This group was known presumptively as the "non-organic group (A)". It was planned to obtain neurological evaluation of members of this group at a later date, if certain obstacles could be overcome. A fourth group of 11 non-organic children was added about a year later and designated as the "non-organic group (B)".

## Results

Table 1 indicates the mean ages, IQ's, and ratios of males to females for the four groups. The N is small in each group, but no attempt was made to select the cases. We simply used every case in the psychologist's files that met the requirements of having a WISC Full Scale IQ of 80 or below.

Table 1. Mean Ages, IQ's, and Ratios of Males to Females

| Group | M | F | Ratio | W-B Scales | Mean IQ | S. D. |
|---|---|---|---|---|---|---|
| Organic (Mean CA 10-2) N = 27 | 21 | 6 | 3.5:1 | Full Scale | 56.34 | 10.14 |
| | | | | Verbal Scale | 62.44 | 11.00 |
| | | | | Performance Scale | 58.00 | 11.91 |
| Suspected (Mean CA 10-6) N = 48 | 32 | 16 | 2.0:1 | Full Scale | 61.33 | 11.15 |
| | | | | Verbal Scale | 65.70 | 10.46 |
| | | | | Performance Scale | 62.94 | 13.32 |
| Non-organic (A) (Mean CA 11-5) N = 29 | 18 | 11 | 1.6:1 | Full Scale | 69.38 | 9.57 |
| | | | | Verbal Scale | 70.76 | 8.89 |
| | | | | Performance Scale | 74.21 | 13.11 |
| Non-organic (B) (Mean CA 10-3) N = 11 | 7 | 4 | 1.7:1 | Full Scale | 69.40 | 8.20 |
| | | | | Verbal Scale | 73.50 | 9.75 |
| | | | | Performance Scale | 73.05 | 12.50 |

Table 1 shows the distribution of sexes in the various groups. Although the ratio of boys to girls in the organic group is somewhat higher than the 2.9:1 ratio that Strauss[11] reports in his exogenous group and the 2.7:1 ratio that Penrose[10] reports in his brain-injured group, the ratio in the non-organic groups compares favorably with Strauss'[11] endogenous group. This excess in our organic group can probably be accounted for by the fact that several of the cases have been diagnosed as cerebral dysplasia which would probably place them in what Strauss[11] refers to as the endogenous group. The factor of sexual differences is of doubtful importance in relation to the main topic of this study, but we felt that it might be of some value or interest to some readers.

Tri-serial correlations[7] were obtained between organicity and intelligence as indicated by Full Scale, Verbal, and Performance IQ's. These were .493, .335, and .502 respectively. These correlations indicate that there is no significant degree of relationship between organicity and intelligence as measured by the Verbal Scale. The limits of confidence for the correlations obtained on the Performance and Full Scales lie between 10 and 20%.

Table 2 shows the degree of relationship between intelligence and organicity as indicated by bi-serial correlations. Although organicity might be considered an either/or condition, bi-serial correlation is warranted since the organicity is inferred from the symptoms presented. The symptoms are actually what are measured or evaluated, and as they may vary in number and severity they may be placed on a continuum. The degree of distinction between the organic and non-organic group is greatest and the distinction between the organic and suspected organic group is least on all three scales. It will also be noted that the degree of distinction is less on the Verbal Scale than either the Full Scale or Performance Scale.

TABLE 2. BI-SERIAL CORRELATIONS INDICATING THE DEGREE OF RELATIONSHIP BETWEEN ORGANICITY AND INTELLIGENCE, AND DEGREE OF SIGNIFICANT DIFFERENCES BETWEEN GROUPS

| Groups | N | Full Scale | | | Verbal Scale | | | Performance Scale | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $r_{bis}$ | $t$ | level conf. | $r_{bis}$ | $t$ | level conf. | $r_{bis}$ | $t$ | level conf. |
| Organic and Non-organic | 56 | .704 | 15.25 | .01 | .470 | 9.72 | .01 | .681 | 16.85 | .01 |
| Suspected and Non-organic | 77 | .449 | 10.58 | .01 | .302 | 6.93 | .01 | .488 | 12.99 | .01 |
| Suspected and Organic | 75 | .276 | 6.30 | .01 | .187 | 4.06 | .01 | .234 | 4.25 | .01 |

It will be noted in Table 2 that there is a significant difference between the mean IQ's of the various groups beyond the 1% level of confidence. It is also significant that the performance scale mean IQ is less than the verbal scale mean IQ for the organic group and the suspected group while the non-organic group shows a performance mean IQ higher than the verbal scale mean IQ. This agrees with Wechsler's[14] observation that organics usually have a lower performance IQ than verbal IQ. Vanderhost, Sloan, and Bensberg[13] reported that high-grade mental defectives matched for CA, MA, and IQ were found to be uniformly higher on performance than verbal tests on both the Wechsler-Bellevue and the Wechsler Intelligence Scale for Children. Sloan and Schneider[12] in another study utilizing the WISC reported that the performance IQ's of mental defectives were found to be consistently higher than verbal IQ's. It would appear that the lower the subject's IQ on the various scales the greater the probability of organicity, and the higher the subject's IQ the smaller the probability of organicity.

In trying to arrive at a characteristic pattern of subtest scores, the deviation of each individual's subtest scores from his own average subtest score was computed. These deviations were then averaged for each subtest and each group, and then rated according to the following criteria as outlined by Wechsler[14] for subjects having total scores beyond the 80-110 limits.

> "For subjects having total scores beyond the limits 80-110, the deviation of the individual's subtests from their mean, defining a significant difference may be roughly obtained by dividing the mean subtest score by 4. Example: Subject A makes a full score of 56. The mean subtest score is accordingly 5.6 points; one fourth of this is 1.4 points. Hence, any of his individual subtest scores which deviate by more than 1.5 points would be significant."

The mean subtest scores for each group were computed along with the necessary deviations required to be significant. Since only one subject in the organic group actually had this type of pattern the writers felt that this was not a characteristic pattern for organics on the WISC.

The writers are of the opinion that at the present time there is no substitute for experience and observation of the qualitative aspects of the subject's test performance regardless of the existence or non-existence of quantitative test characteristics. In other words the quantitative findings are another aspect of the total picture. For as Wechsler[14] indicates, the organic may or may not follow the general pattern.

It might be stated here that these results are not surprising since the literature indicates some rather discouraging results using the W-B, Form I, Scale for the detection of organicity[1, 4, 5, 6, 8, 15].

In following up the "non-organic group (A)" and the "suspected group", almost a year's time elapsed. Of the non-organic group three received neurological examinations, and were found to be non-organic. Seven of the suspected group have been studied neurologically, and three of the seven were diagnosed as non-organic. Five additional children not included in the original groups were examined and found to be non-organic, resulting in a group of eleven medically diagnosed non-organics, designated as "non-organic group (B)". This group compared very favorably with the original non-organic group as the Mean Full Scale IQ's are almost identical as indicated in Table 1, and lends weight to the argument that the possibility of organicity increases significantly as the IQ drops below the borderline feeble-minded range.

The authors feel that the following factors should be taken into consideration in any further study of this type:

1. Mentally handicapped organics have such low scores that there is less chance for variation in the subtest scores.
2. At the lower age levels the subject may receive a scaled score even though his raw score is zero.
3. Different types of organics may offset each other's scores as some types may do better than others on certain subtests and vice versa.
4. Organic children differ from organic adults in that they have never developed intellectually to the extent that the adults have whose organicity occurs later in life. Therefore the subtests which hold up for adults, such as vocabulary, to produce scatter do not hold up for children, as all learning is new learning for them.

## Summary

1. Thirty-one of forty-two children with IQ's below 80 who because of test data were suspected of suffering from organic brain disease were found to have confirmatory signs on neurological examinations.
2. Organics tend to score lower on the WISC Full Scale than non-organics.
3. Organics tend to do more poorly on the WISC Performance and Full Scales than on the WISC Verbal Scale.
4. The possibility of organicity increases considerably as the IQ drops below the borderline feeble-minded range, i.e., below 70-80 on the WISC.
5. The WISC does not show a characteristic pattern of subtest scores for organics as a group that is representative of the individual test patterns using the method outlined by Wechsler[14].
6. Further research is indicated to determine whether or not different types of organics will show specific test patterns which are significant in differentiating organics from non-organics.

## References

1. Anderson, A. Lloyd. The effect of laterality localization of brain damage on Wechsler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1950, 6, 191-194.
2. Anderson, H. H. and Anderson, G. L. *An introduction to Projective Techniques.* New York: Prentice Hall, 1951.

3. BURTON, A. AND HARRIS, R. E. *Case Histories in Clinical and Abnormal Psychology.* New York: Harper, 1947.
4. COHEN, JACOB. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, 47, 359-365.
5. FOX, CHARLOTTE AND BIRREN, JAMES E. Intellectual deterioration in the aged: Agreement between the Wechsler-Bellevue and the Babcock-Levy. *J. consult. Psychol.*, 1950, 14, 305-310.
6. HOWARD, ALVIN R. Diagnostic value of the Wechsler Memory Scale with selected groups of Institutionalized patients. *J. consult. Psychol.*, 1950, 14, 376-380.
7. JASPEN, NATHAN. Serial Correlation. *Psychometrika*, 1946, 2, 23-30.
8. JASTAK, JOSEPH. Ranking Bellevue subtest scores for diagnostic purposes. *J. consult. Psychol.*, 1953, 17, 403-410.
9. LINDQUIST, E. F. *Statistical Analysis in Educational Research.* Cambridge: Riverside Press, 1940.
10. PENROSE, L. S. *A Clinical and Genetic Study of 1280 Cases of Mental Defect.* His Majesty's Stat. Off., London: 1938.
11. STRAUSS, A. A. AND LEHTINEN, L. E. *Psychopathology and Education of the Brain-Injured Child.* New York: Grune and Stratton, 1947.
12. SLOAN, WILLIAM AND SCHNEIDER, BERNARD. A study of the Wechsler Intelligence Scale for Children with mental defectives. *Amer. J. Mental Deficiency*, 1951, 55, 573-575.
13. VANDERHOST, LEONETTE, SLOAN, W., AND BENSBERG, GERARD J., JR. Performance of mental defectives on the Wechsler-Bellevue and the Wechsler Intelligence Scale for Children. *Amer. J. Mental Deficiency*, 1953, 57, 481-483.
14. WECHSLER, D. *The Measurement of Adult Intelligence.* Baltimore: Williams & Wilkins, 1944.
15. WITTENBORN, J. R. AND HOLZBERG, J. D. The Wechsler-Bellevue and descriptive diagnosis. *J. consult. Psychol.*, 1951, 15, 325-329.

# TEACHERS' JUDGMENTS OF THE POPULARITY OF CHILDREN FROM THEIR HUMAN FIGURE DRAWINGS[1]

ALEXANDER TOLOR

*Columbia-Presbyterian Medical Center, N. Y. C.*

## PROBLEM

In a previously conducted study [3] it was demonstrated that experienced clinical psychologists are, in general, successful in discriminating between the more and less popular children in classes by globally evaluating their human figure drawings. The present study was designed to determine whether elementary school teachers, who have had no formal training in the analysis of projective techniques, can also make this type of differentiation. More specifically, this investigation concerns itself with the following problems: (1) Is the difference in training and experience between teachers and psychologists related to the level of accuracy achieved in discriminating between the drawings of more and less popular children? (2) Do psychologists or teachers agree among each other to a greater degree? (3) Do representatives of these two groups consider similar personality dimensions to be significantly related to popularity? and (4) Do teachers and psychologists utilize the same formal characteristics of graphic productions in forming their impressions?

If it is true that clinical psychologists have developed valid and specialized methods for analyzing projective drawings which they are able to apply to the evaluation of one major aspect of social adjustment, i. e., to the assessment of the degree of acceptance or rejection by peers, then they would be expected to be considerably more accurate in their judgments and to show a greater degree of agreement in their

[1]The author wishes to express his appreciation to the teachers of P. S. 75, Brooklyn, N. Y. who cooperated in this study by serving as judges.

selections than members of lay groups. However, in case psychologists actually must rely on techniques which are not unique to themselves, but which are in practice commonly, although perhaps unconsciously, utilized by psychologically less sophisticated individuals, then one would expect psychologists and teachers to be approximately equal in accuracy of ratings and in the degree of agreement in their selections. Thus a comparison of performance in this area of teachers and psychologists assumes considerable theoretical importance.

## Method

One hundred and thirty six white children in three fifth grade classes and in one fourth grade section attending the same metropolitan public school were employed. The fifth grade subjects ranged in age from 10 years 1 month to 12 years 8 months (mean 10 yrs. 10 mos.); in IQ they ranged from 70 to 125 (mean 94.7). The fourth grade children had an age range of 9 years 1 month to 10 years 6 months (mean 9 yrs. 7 mos.); their IQ's ranged from 70 to 135 (mean 103.3). The entire group contained approximately as many boys as girls.

Sociometric status of the children in each class was determined by requiring each child to select three associates whom he would like to invite to his home for dinner on a specific night. The sociometric investigation was conducted near the end of the school year at a time when it may be presumed that the children had had ample opportunity to become well-acquainted with one another. Relative popularity was defined in terms of the frequency with which any one child was selected by his classmates. That this criterion of popularity is a fairly reliable index is indicated by the product-moment correlations for the four sections which ranged from .71 to .76 with a one week test-retest interval.

Within two days of the original sociometric analysis all children produced two drawings in class; one in response to the instruction that they draw a person and the other in response to the request that they draw a person of the sex opposite to the one initially executed. The children labeled the sex of their drawings, and clearly marked the order in which these were produced.

The human figure drawings of all the children in each section were then presented to five teachers who were not acquainted with these subjects. As a further precaution against extraneous cues in those cases where children had included their names on the paper, this information was deliberately deleted by the experimenter. The teachers, all with a minimum of 1½ years of teaching experience, were instructed to select in each class the five sets of drawings which they thought were the product of the most popular and the five sets of drawings representative of the least popular children. The manner in which the criterion of popularity had been established was clearly explained to the judges.

All of the teachers were further requested to supply information pertaining to the manner in which they arrived at their decisions. They indicated the specific personality characteristics which they considered to be significantly related to the trait of popularity and listed the essential features contained in the drawings which they had utilized in forming their impressions. Since in the previously cited experiment [3] similar information had been obtained from a group of clinical psychologists subsequent to their judgments of drawings, a basis for comparison was readily available.

## Results

Inspection of Table 1 reveals that when the judgments of the five educators are combined for each class, the teachers are able to differentiate between the more and less popular children to a statistically significant degree in only one of the four sections. This finding stands in sharp contrast to that of the five psychologists who were successful in making accurate discriminations in three of the four sections (the exception being the fourth grade class).

TABLE 1. SUMMARY OF THE COMBINED ABILITY OF FIVE TEACHERS TO DIFFERENTIATE IN EACH OF FOUR SECTIONS BETWEEN MORE AND LESS POPULAR CHILDREN. THE FREQUENCIES REPRESENT SOCIOMETRIC RATINGS.

| Section | Popular | Unpopular | Chi-square | P |
|---|---|---|---|---|
| I | 84 | 41 | 14.80 | <.01 |
| II | 93 | 87 | .20 | —— |
| III | 82 | 59 | 2.76 | —— |
| IV | 68 | 60 | .50 | —— |

Table 2 presents data showing that three of the five teachers, when their ratings of drawings for the four sections are combined, discriminated to a statistically significant degree between the more popular and the less popular students. This compares rather favorably to the finding of the previous study in which four out of five psychologists were successful in differentiating between the extremes in popularity.

TABLE 2. SUMMARY OF THE ABILITY OF EACH TEACHER TO DISCRIMINATE BETWEEN THE MORE AND LESS POPULAR STUDENTS WHEN THE FOUR SECTIONS ARE COMBINED. THE FREQUENCIES REPRESENT SOCIOMETRIC RATINGS.

| Teacher | Popular | Unpopular | Chi-square | P |
|---|---|---|---|---|
| A | 68 | 43 | 5.63 | .02 |
| B | 67 | 65 | .04 | —— |
| C | 58 | 62 | .14 | —— |
| D | 73 | 41 | 8.98 | .01 |
| E | 61 | 36 | 6.44 | .02 |

Analysis of the data also permitted answering the question of whether these teachers can make the more difficult, but finer discriminations between children who are average in popularity and those who fall within one of the extremes. It is conceivable that judges who attain some measure of success in differentiating between the most and the least popular students may remain unsuccessful in discriminating between children who differ to a lesser degree in popularity. In this connection, reference should be made to the five clinicians who were generally less successful in the latter type of task than in the former since only three of the five psychologists achieved statistical significance, one for his popular selections, two in selecting drawings of unpopular children as compared to the average. When Chi-square tests are performed for teachers' ratings, it is found that three teachers can differentiate to a statistically significant degree between children who are average in popularity and those who are least popular. None is successful, however, in differentiating between children who are average in popularity and those who are most popular.

Another problem worthy of attention pertains to the degree of agreement among raters in their selection of drawings which were presumed to reflect different levels of popularity. When the number of drawings concerning which three or more teachers agreed is compared to the total number of selections made, 13.9% agreement is found for the group considered to be more popular and 14.3% for the group considered to be less popular. Since the corresponding figures for psychologists were 24.1% and 34.8% respectively, it is evident that these teachers agreed with one another to a considerably lesser extent than the psychologists. A more detailed breakdown of the percent of agreement in each section is presented in Table 3. In all cases except one these figures are lower than those for psychologists.

That this group of teachers and the psychologists agreed between themselves to a greater degree about the drawings of the less popular children than those of the more popular children is indicated by the fact that in 28% of the former cases and in only 13.6% of the latter is there overlapping between the selections of three or more raters in the two groups.

TABLE 3. AGREEMENT OF TEACHERS

| Section | Percent Agreement | |
|---|---|---|
| | More popular | Less popular |
| I | 20.0 | 21.4 |
| II | 11.8 | 12.5 |
| III | 11.1 | 13.3 |
| IV | 13.3 | 11.1 |

In order to investigate the possibility that these teachers inadvertently selected drawings produced by more intelligent students, when they were requested to choose the work of more popular children, and drawings made by less intelligent children, in selecting the work of less popular children, mean IQ's were computed for the two groups. It was found that the average IQ of subjects assigned to the more popular group by three or more teachers is 105 whereas the average IQ of subjects assigned to the less popular group is only 84, a difference which is statistically significant. This same relationship of intellectual differences to differences in popularity was also noted in the judgments of psychologists, although to a far lesser degree.

*Qualitative Findings.* In an attempt to ascertain the manner in which the judges arrived at their decisions concerning the child's level of popularity, and to determine whether their methods differed from that of psychologists, teachers were requested to furnish additional information. They indicated what personality factors they thought would be associated with a child's popularity standing within his group and listed the characteristics of drawings upon which they relied in formulating ideas. Although these introspective statements were not subjected to quantitative analysis, their inspection permits the following tentative conclusions: (1) Teachers often have some vague notion of what features of the drawings yield significant personality clues but they do not seem to know precisely how these formal characteristics should be interpreted. (2) These teachers seem to show a greater diversity of views than psychologists concerning the manner in which the personality trait of popularity is thought to fit into a broader concept of adjustment.

## DISCUSSION

Recent studies, such as the one by Holsopple and Phelan [1], have focussed attention on the efficacy of clinicians in utilizing projective techniques. Since even psychologists seem to perform at varying levels of efficiency, it was thought that a comparative investigation of the skill of clinicians and teachers would shed some light on the ability of individuals having different types of experiences and training to interpret projective material. For this purpose, human figure drawings were selected since their analysis does not require the mastering of any complicated scoring technique and may therefore be submitted directly to members of lay groups for evaluation. Although Karen Machover [2] does not refer directly to the problem of ascertaining the level of group acceptance or rejection from projective drawings, it may be assumed that the interrelationship of drawing characteristics, since it has significant implications for level of adjustment, can be used for this purpose.

All judges were requested to make their decisions on the basis of global impressions rather than specific drawing characteristics. Since most teachers have had at least some opportunity to see children's graphic productions, it is likely that this approach was fairer to the teachers than a more atomistic one would have been.

The results are obviously limited by the small number of teachers used in the sample and by the equally small group of psychologists with whom comparisons were made. Keeping in mind the hazards involved in generalizing to other groups, the findings suggest that there are very real differences between these teachers and psychologists in judging the relative popularity of children on the basis of their human

figure drawings. Psychologists as a group tend to be superior to the teachers in making this type of differentiation, although individual teachers compare rather favorably to psychologists. Unlike the clinicians, who are approximately equal in ability to detect the drawings of very unpopular and very popular children, the teachers seem to have considerably more success in identifying the work of the less popular subjects. In general, teachers agree less among each other in their choices than do clinicians who appear to rely more on a specific theoretical framework. The teachers, on the other hand, depend very much on those differences in drawings which are related to differences in intellectual capacity.

## Summary

In a recent study[3], in which clinical psychologists judged the human figure drawings of elementary school children, it was found that differentiation between students who were more popular and those who were less popular was possible with considerable accuracy. The present investigation, in which the same projective drawings are used, was designed to determine how a group of teachers compared to psychologists in making similar evaluations. The relative popularity of the participating children was determined through sociometric studies. The drawings of human figures of both sexes, which were produced by 136 children, were submitted to five elementary school teachers who were requested to select the five sets in each class which were representative of the most popular children, and the five sets drawn by the least popular children. The teachers also supplied information pertaining to how they arrived at their decisions.

The data indicated that these teachers, as a group, are far less successful than psychologists in making this type of differentiation although some of the teachers, individually, attain surprisingly good levels of discrimination, especially in detecting the drawings made by children who tend to be rejected by their peers. Other findings were that these teachers show less agreement than psychologists in their selections, that their choices are closely related to intellectual differences of the children executing these drawings, and that they are less aware of the factors influencing their decisions than psychologists. Within the limitations of the sample and methods used, these results suggest that the judgment of children's popularity from their human figure drawings requires highly specialized training and experience which is more compatible with that of clinical psychologists than elementary school teachers.

## References

1. Holsopple, J. Q. and Phelan, J. G. The skills of clinicians in analysis of projective tests. *J. clin. Psychol.*, 1954, *10*, 307-320.
2. Machover, Karen. *Personality projection in the drawing of the human figure.* (3rd Ed.) Springfield, Ill.: C. C. Thomas, 1953.
3. Tolor, A. and Tolor, Belle. The judgment of children's popularity from their human figure drawings. *J. proj. Tech.*, in press.

# CONSISTENCY OF REPRODUCTIONS OF BENDER-GESTALT DESIGNS[1]

MARION WHITE MC PHERSON
*The Neuropsychiatric Clinic of St. Louis, Missouri*

LORETTA A. PEPIN
*The Public Schools of Detroit, Michigan*

## Problem

The present paper is a report of the similarity of reproductions of the Bender-Gestalt designs when two different methods of motor manipulation are employed. If the responses obtained by the two methods agree, it is rational to conclude that the reproductions of the stimuli are relatively independent of the particular motor techniques with which they were executed. If the responses do not agree, it can be inferred that motor factors obliterate perceptual, covert contributors to these particular responses.

## Subjects

The subjects were 20 male and 12 female students, all the members of a senior college course in personality measurement. Bias resulting from a volunteer sample was eliminated but the group is highly selective in terms of factors that determine enrollment in such a course. At the time of observation the subjects had not received formal training in projective testing but undoubtedly had some general familiarity with the concepts and techniques.

## Procedure

The subjects were asked to duplicate the Bender Gestalt designs on paper and by placing pieces of felt on a felt board. The order in which the felt and paper reproductions were obtained was counterbalanced. Customary clinical procedure characterized the obtaining of the paper reproductions. In the felt board situation only Designs A, 1, 2, 3, and 5 were utilized. These particular designs appeared to be the most adaptable to felt and to be adequate for an initial investigation. Each subject was requested to duplicate the designs by placing smaller pieces of felt on a board 34½″ x 46½″. In front of this were 950 individual pieces of felt separated according to color, size, and shape. There were seven different colors as well as black and white circles and ovals, both with and without solid centers, and with diameters varying from 5/8″ to 4″ as well as 2″ squares. The chromatic pieces were available to serve the purposes of an allied investigation of the consistency of color use. The various hues which the subjects used are not of immediate interest and no description of them will be included in this paper.

Time expediency did not permit the scoring of the designs as each subject completed them so a plastic sheet marked into squares was placed over the felt board to prevent accidental displacement of the designs. Sheets of transparent paper the same size as the squares were then placed over the appropriate squares and the individual pieces of felt as placed by the subjects traced. This left a series of lettered papers which could be used later to reconstruct the designs.

Scoring items comparable to the two media were located from the scoring suggestions offered by Billingslea[(1)], Hutt[(3)], Pascal and Suttell[(4)], and by empirical inspection of the present data. Because the actual size of both the component designs and the backgrounds varied in the two situations, linear and spatial measurements were not expedient and the designs were rated as like or deviant from the standard. Inspection of the actual reproductions revealed varying degrees of deviation. For example, blunting of Design 3 was never extreme so this item was rated

---

[1]When this research was carried out the senior author was Assistant Professor of Psychology in the College of Liberal Arts of Wayne University, Detroit, Michigan.

The authors wish to express their appreciation to Mr. Frederick G. Seaver of the American Felt Company who made available the felt used in this investigation. He was most cooperative in supplying material of specified size and color.

on a 2-point scale and each reproduction scored as either correct, 0, or incorrect, 1. Errors on some items were found to be markedly deviant, 2, or extremely so, 3. The actual data indicated that description of dissimilarity from the standard could be most accurately achieved by the use of rating scales of four different lengths.

Table 1 lists the 40 items that were rated, the characteristics by which each was judged, and the number of steps in the rating scale. In some instances a subject omitted a design on either felt or paper. In order to maintain a constant $n$ an omission was rated as maximally deviant for the medium on which it was missing. The actual ratings were assigned by inspecting all felt duplications and then all paper reproductions.

TABLE 1. ITEMS THAT WERE SCORED

| Design | Title | Characteristics |
|---|---|---|
| | | *2-point scales* (01) |
| A | Contact | Overlap or separation of the 2 component parts |
| A, 1, 2, 3, 5 | Consistency | Shape of ½ or more of the components varies |
| 1 | Double Row | Design reproduced on 2 lines |
| 2 | Two Lines | Base of design on more than one level |
| 3 | Blunting | Point of arrow flattened |
| 5 | Extension rotation | Extension rotated or direction reversed |
| Configuration | Compression | All designs in half or less of area |
| | | *3-point scales* (02) |
| 1, 2, 3 | Spacing (h) | Lack of uniform spacing-horizontal axis |
| 3 | Spacing (v) | Lack of uniform spacing—vertical axis |
| 1,2 | Wavy Line | Base Line not straight |
| 2 | Deviation Slant | Column slant deviates at least twice |
| 2 | Alignment | Column alignment deviates at least twice |
| 3 | Disparity | Angles to horizontal axis unequal |
| 3 | Axis Straightness | Horizontal axis not straight |
| 3 | Axis Ratio | Vertical width equal or greater than horizontal |
| 5 | Dot Intersection | Extension intersects curve at a dot |
| 5 | Asymmetry-Length | Sides of curve unequal in length |
| 5 | Asymmetry-Contour | Contour of sides of curve deviates |
| | | *4-point scales* (03) |
| A, 1, 2, 3, 5 | Rotation | Design rotated |
| Configuration | Order | Order of reproduction deviates from standard |
| Configuration | Relative Size | Variation in relative size of designs |
| | | *Continuum scales* |
| A, 1, 2, 3, 5* | Parts of Design | Number of component parts |
| 2 | Touching | Number of places components touch |
| Configuration | Overlap | Number of designs projecting into others |

*In Design 5 a count was made of the number of component parts in both the curve and extension of the design.

In order to describe similarity between the two media the scores were combined. For example, a score of 02 indicates an accurate reproduction on felt but a deviant one on paper. The four rating scales could produce varying combinations. For example, the 2-point scales could effect only combinations of 00, 01, or 11, but the 3-point scales could result in tallies of 00, 01, 02, 11, 12, or 22. Each combination was categorized under one of the following five divisions of agreement:

*Total agreement—correct* ($TA_c$): the absence of deviation from the standard in both situations (00 combinations on all scales).

*Total agreement—incorrect* ($TA_i$): the presence of common error in both situations (2-point scales, 11; 3-point scales, 11 or 22; 4-point scales, 11, 22 or 33; and, continuum scales, an equal number of stimuli correct in both media).

*Intermediate agreement*—($IA_a$): correctness in one situation and partial error in the other or errors in both media that were no more than one rated step from one another (2-point scales, none; 3-point scales, 01 or 12; 4-point scales, 01, 12 or 23; and, continuum scales, 0 in one medium and at least ½ the correct number of stimuli in the other).

*Intermediate agreement—($IA_b$):* correctness in one situation and extreme error in the other or only partial correctness in the two situations. In both instances the ratings were more than one step removed from one another (2-point scales, none; 3-point scales, none; 4-point scales, 02 or 13; and, continuum scales, deviation of at least ½ the correct number of stimuli in both media).

*Total disagreement—(TD):* correctness in one situation and total incorrectness in the other (2-point scales, 01; 3-point scales, 02; 4-point scales, 03; and, continuum scales, 0 on one medium and total incorrectness on the other).

## Results

Table 2 presents for each scored item the number of cases in each of the five divisions of agreement. It also lists the sum of these occurrences, the possible frequency of each category of agreement for this group of 32 subjects scored on 40 items, and the per cent the actual incidence was of the possible.

Table 2. Incidence of Varying Amounts of Agreement in the 40 Scored Items

| Design | | Item | $TA_c$ | $TA_i$ | $IA_a$ | $IA_b$ | TD |
|---|---|---|---|---|---|---|---|
| 2-pt. | A | Contact | 14 | 4 | x* | x | 14 |
| | A | Consistency | 30 | 1 | x | x | 1 |
| | 1 | Consistency | 27 | 1 | x | x | 4 |
| | 2 | Consistency | 7 | 6 | x | x | 19 |
| | 3 | Consistency | 30 | 0 | x | x | 2 |
| | 5 | Consistency | 28 | 1 | x | x | 3 |
| | 1 | Double Row | 30 | 0 | x | x | 2 |
| | 2 | Two Lines | 31 | 0 | x | x | 1 |
| | 3 | Blunting | 29 | 0 | x | x | 3 |
| | 5 | Extension Rotation | 31 | 1 | x | x | 0 |
| | Conf. | Compression | 29 | 0 | x | x | 3 |
| 3-pt. | 1 | Spacing (h) | 26 | 0 | 4 | x | 2 |
| | 2 | Spacing (h) | 19 | 4 | 7 | x | 2 |
| | 3 | Spacing (h) | 25 | 3 | 4 | x | 0 |
| | 3 | Spacing (w) | 17 | 3 | 11 | x | 1 |
| | 1 | Wavy Line | 16 | 3 | 12 | x | 1 |
| | 2 | Wavy Line | 8 | 9 | 13 | x | 2 |
| | 2 | Deviation Slant | 26 | 1 | 4 | x | 1 |
| | 2 | Alignment | 19 | 1 | 11 | x | 1 |
| | 3 | Disparity | 13 | 2 | 15 | x | 2 |
| | 3 | Axis Straightness | 16 | 0 | 16 | x | 0 |
| | 3 | Axis Ratio | 15 | 3 | 14 | x | 0 |
| | 5 | Dot Intersection | 8 | 14 | 5 | x | 5 |
| | 5 | Asymmetry—Length | 25 | 2 | 4 | x | 1 |
| | 5 | Asymmetry—Contour | 30 | 0 | 1 | x | 1 |
| 4-pt. | A | Rotation | 26 | 0 | 1 | 2 | 3 |
| | 1 | Rotation | 28 | 1 | 1 | 0 | 2 |
| | 2 | Rotation | 30 | 0 | 1 | 0 | 1 |
| | 3 | Rotation | 27 | 2 | 1 | 1 | 1 |
| | 5 | Rotation | 27 | 1 | 0 | 2 | 2 |
| | Conf. | Order | 15 | 1 | 8 | 3 | 5 |
| | Conf. | Relative Size | 4 | 12 | 12 | 4 | 0 |
| Cont. | A | Parts of Design | 31 | 0 | 1 | 0 | 0 |
| | 1 | Parts of Design | 28 | 0 | 3 | 1 | 0 |
| | 2 | Parts of Design | 28 | 0 | 4 | 0 | 0 |
| | 3 | Parts of Design | 28 | 0 | 4 | 0 | 0 |
| | 5 | Parts of Design (c) | 23 | 1 | 8 | 0 | 0 |
| | 5 | Parts of Design (t) | 22 | 2 | 7 | 1 | 0 |
| | 2 | Touching | 31 | 0 | 1 | 0 | 0 |
| | Conf. | Overlap | 12 | 1 | 18 | 1 | 0 |
| | | Possible Occurrence | 1280 | 1280 | 928 | 480 | 1280 |
| | | Actual Occurrence—sum | 909 | 80 | 191 | 15 | 85 |
| | | Actual Occurrence—% | 71% | 6% | 21% | 3% | 7% |

*No tally possible at points marked with (x).

This table reveals that totally correct responses in both media occur 71% of their theoretical possibility. Responses that are similarly deviant in both media appear 6% of the possible frequency. Both of these indicate consistency of behavior and when added reveal that such consistency happens 77% of the theoretical frequency. Total disagreement actually appears only 7% of its theoretical incidence. Partial discrepancy, IAa, is more frequent (21%) than is more extreme deviation, IAb (3%). The discrepancies between total agreement, both correct and incorrect, and the frequencies in the other categories of agreement are so great that the significance of the results appears assured without the formal application of the customary statistical tests of significance. Such data make it rational to infer that the Bender-Gestalt is not predominantly influenced by lack of motor control *per se*. These data thus provide the groundwork for additional research designed to estimate the points at which the definiteness of reality and the amount of clinical disturbance alter this pattern of perceptual consistency.

Since such a pattern of predominantly correct agreement is to be expected from clinically normal individuals, violation of it or incidences of disagreement are of special interest. Disagreement could be promoted by the incomparability of the two media or by lack of perceptual discrimination on the part of the subjects as well as by inconsistency in motor behavior. Since all but one subject contributed extreme scores the final tally does not appear to be biased by individual idiosyncracies. Because the ratings were made serially for each design the results are not apt to be influenced by concentrated familiarity with individual protocols.

Analysis of the inconsistencies revealed that in two scoring points discrepancies from the standard were much more frequent on one medium. Such results indicate that the present data are in these cases biased by a methodological artifact of the incomparability of the two media. Removal of these scoring items would increase the amount of agreement above the tally of 77%. Errors on three scoring points were equally divided. Goebtz[2] reports that some deviations were more characteristic of normals than of neurotics. Inconsistencies in the choice of the component parts of Design 2 were equal in both media to the number of correct responses. The cards purchased for this research presented these parts as oval in shape. Those reproduced in Pascal and Suttel[4] are circular and are referred to as such both by these authors and by Goebtz[2]. If the "standard" stimuli vary this much, it is not surprising to find the present sample demonstrating confusion.

## Summary

In this study 32 subjects duplicated all the Bender-Gestalt designs on paper and five of them by placing pieces of felt on a felt board. Their performances were rated as to degree of similarity to the standard stimuli on both media and the ratings obtained on the two compared. The results indicate that uniformity actually occurs a minimum of 77% of its theoretical possibility and that total disagreement or extreme disagreement occurs only 6% of the time. These data are interpreted as indicating that the reproduction of the Bender-Gestalt figures is more influenced by covert perceptual responses than by motor techniques.

## References

1. Billingslea, F. Y. The Bender-Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1948, 4,1-27.
2. Goebtz, Wallace. A quantification, standardization, and validation of the Bender-Gestalt test on normal and neurotic adults. *Psychol. Monog.*, 1953, 67, No. 6, pp. 28.
3. Hutt, M. L. *A tentative guide for the administration and interpretation of the Bender-Gestalt test.* 1945, U. S. Army Adjutant General's School.
4. Pascal, G. R. and Suttell, B. J. *The Bender-Gestalt test.* New York: Grune and Stratton, 1951.

# INTER-RATER AGREEMENT IN JUDGING STUDENT ADJUSTMENT FROM PROJECTIVE TESTS

IJA N. KORNER AND DALE WESTWOOD

*University of Utah College of Medicine*

## Problem

It is very difficult to get at inter-rater reliability of judgments based on projective tests. If stress is placed upon objectivity by means of atomistic division of the protocol into test details, test signs, or scoring categories, the clinician is handicapped because he prefers the global, configural inferences which the theoretical foundations of the test require. On the other hand, if complex inferential statements about personality are used to assess reliability, the problem is complicated because each psychologist may select different variables upon which to base his interpretations. As a compromise between the atomistic and the configural approaches to projective test reliability, we used the global variable of personal adjustment. The task of each judge was to sort the projective test protocols into three equal piles with the guiding precept being only: "Is the subject's adjustment good, average, or poor?"

We assumed that level of personal adjustment is a continuous, normally distributed variable by means of which inter-rater correlations could be obtained. Albee and Hamlin[1] pointed out that adjustment may be a multi-dimensional variable. Likewise the variable of level of intellectual functioning (or IQ) is multi-dimensional.

Numerous studies have appeared which are concerned with inter-rater reliability with respect to one projective test. It was our intention to broaden the approach to projective test reliability by also determining the degree of inter-judge agreement across two tests.

Figure drawings and the Rorschach seem to force different variables into the focus of the evaluator. For example, a patient's figure drawings may impel an evaluation of impulse level, while his Rorschach protocol points up the capacity to integrate both fantasy and reality. As incorporated in a report, both inferences are complementary. Would both tests used independently by the same judges lead to similar evaluation of level of personal adjustment?

## Procedure

Three clinical psychologists independently rated the human figure drawings and then the group Rorschachs of 96 freshman college students enrolled in an introductory psychology course.

The directions to raters were the same for both tests: "Rate each test protocol in terms of level of personal adjustment. Sort the protocols into three equal piles. In the first pile place the tests of those subjects considered best adjusted. The second pile is for those of moderately successful adjustment, and the third pile is for those with the least successful adjustment pattern. Use your concept of adjustment and your criteria for judgment."

The raters were senior members of the Department of Psychiatry, University of Utah, who have been co-workers for over four years. The Rorschach as well as the figure drawings are routinely included in the approximately 400 test batteries given by the staff per year.

In order to correlate the ratings of the judges the protocols in each pile were assigned linear scores on the assumption that adjustment is normally distributed. Those protocols in the best adjusted group were given scores of 69, those in the middle group scores of 50, and those in the least successful adjustment group scores of 31. Product-moment inter-correlations for the three judges were computed for the figure drawings and also for the Rorschach protocols. Correlations between the

ratings on the Rorschach and the figure drawings were computed for each judge and for the pooled ratings of the three judges.

In addition to the projective tests the Otis Self-Administering Test of Mental Ability (Form A) was given to each student. The students' grade-point-averages for the term during which the tests were administered were available.

## Results

For the figure drawings the correlations between raters are statistically stable beyond the .01 level. For the group Rorschach two of the three inter-raters correlations are statistically stable at the .01 level, but are relatively low values. Since two of the *r*s are very much higher than the third it appears that two judges placed emphasis on different test variables with the third judge occupying an intermediate position.

When the ratings for the figure drawings are correlated with those for the Rorschach the values are low, but statistically stable. The inter-test *r*s for two of the raters are stable beyond the .01 level, while that for the third rater is stable at the .05 level. The inter-test correlation using the pooled ratings of the three judges is .35 which is stable beyond the .01 level.

Table 1. Correlations Between Judges' Ratings of Adjustment Based on the Figure Drawings*

| Judge | C | A |
|---|---|---|
| A | .68 | |
| B | .68 | .59 |

*Correlations of .27 are statistically stable at the .01 level (df = 94).

Table 2. Correlations Between Judges' Ratings of Adjustment Based on the Rorschach*

| Judge | C | A |
|---|---|---|
| A | .44 | |
| B | .41 | .09 |

*Correlations of .27 and .20 are statistically stable at the .01 and .05 levels respectively (df = 94).

Table 3. Correlations Between Adjustment Ratings of the Figure Drawings Versus the Rorschach*

| Judge | A | B | C |
|---|---|---|---|
| A | .20 | | |
| B | | .27 | |
| C | | | .35 |

*Correlations of .27 and .20 are statistically stable at the .01 and .05 level respectively (df = 94).

The *r*s between the grade-point-averages (for the term during which the tests were administered) and the pooled ratings are negligible and statistically unstable for both projective tests. Also, for both tests, the *r*s between the Otis and the pooled ratings are negligible and unstable.

TABLE 4. INTERCORRELATION OF POOLED RATINGS, GRADE POINT AVERAGES AND OTIS IQS*

| Variable | Figure Drawing | Rorschach | G. P. A. |
|---|---|---|---|
| Rorschach | .35 | | |
| G. P. A. | .10 | .06 | |
| Otis | .09 | −.08 | .48 |

*Correlations of .27 and .20 are statistically stable at the .01 and .05 level respectively (df = 94).

## DISCUSSION

Our results show considerable agreement between raters. It is relatively high for the figure drawings and relatively low for the Rorschach.

On the drawings the raters show relatively high agreement in spite of the absence of common training, and conflicted attitudes on the part of the three raters concerning the usefulness of figure drawings as clinical instruments. Rater B has had considerable formal training and experience in the clinical evaluation of drawings per se. Rater A is primarily autodidactic and uses drawing intepretations only as an adjunct to other clinical tests. Rater C was trained to interpret drawings in conjunction with a battery of tests. The raters are much closer in their opinions, prejudices, and uses of the Rorschach. Rater B was Klopfer-trained, Rater A was a student of Rapaport, while Rater C was primarily self-trained.

It may be argued that the greater agreement in favor of the figure drawings is due to the fact that the rater is presented with two unified patterns of stimulation (one for the male figure and one for the female figure), while the Rorschach presents ten patterns—one pattern for the responses to each blot—and the complexity results in greater disparity among the raters. This notion could be checked in future investigations by using the responses to only one or two blots as the basis for judgment. One can also maintain that Rorschach interpreters are schooled in unifying ten impressions into an overall evaluation, while figure drawing interpretation demands an analysis of sub-parts which often contradicts the first, unified impression of the drawings.

In response to questioning, each rater maintained that he attempted in both tests to obtain general impressions, but insisted that the factors entering into his Rorschach impressions were very different from those for the drawings.

As one of the raters pointed out: "In rating the drawings I was much less concerned with picking up cues than I was with the absence of pathological characteristics. If a subject's figure drawings had no transparencies, no gross distortions of proportion, etc., I just automatically failed to consider it for the least successful adjustment group. On the other hand, using the Rorschach I was concerned with irregularities. When a person added unusual adjectives to a usual concept or produced usual concepts in an unusual sequence, I automatically excluded this record from the best adjusted group."

Granted observations such as the foregoing are anecdotal they add to our impression that the raters used different criteria, depending upon which of the two tests is under evaluation rather than similar criteria under the handicap of one or two versus ten impressions.

Considering that relatively normal subjects were used, we were surprised that there was as much inter-rater agreement as was found, since the three raters had never explicitly discussed their individual criteria for personal adjustment, or their methods of interpreting either of the two projective tests. The fact that the raters have worked together for over four years undoubtedly weighs heavily in favor of unconscious common learning; nevertheless, in the light of the differences in psycho-

logical background and preparation the considerable degree of agreement appears noteworthy.

A future step in experimentation might be an attempt to separate out the elements producing agreement from those producing disagreement. Is an esoteric dimension such as the *style* of Rorschach responses as well as the *style* of the figure drawings a common dimension for evaluation? Could a detail like the treatment of the neck on a drawing, or an adjective added to a Rorschach percept be the source of marked disagreement among raters?

The raters were apparently not affected by the variable of overall level of intellectual functioning since the Otis IQs showed almost zero correlation with the adjustment ratings. The negligible $r$ of .09 between an intelligence test and level of adjustment inferred from figure drawings is similar to that found by Albee and Hamlin[1]. They found that "The linear correlation between Wechsler-Bellevue Full Scale IQs and the rank-adjustments determined by the case-records and by the judgments of drawings were both negligible and insignificant"[1, p. 391].

No objective evidence is available that the raters were not influenced by the degree of artistic excellence of the drawings. The study of Whitmyre in which he concluded that "Clinical psychologists utilized the artistic merits of the human figure drawings to a large extent in their adjustment evaluation of them"[2, p. 424] was known to each rater and has been repeatedly discussed among them.

A cursory examination of the figure drawings as distributed by the pooled ratings showed a tendency to exclude highly artistic drawings from the "best adjusted" group. This observation, along with others of a similar nature, raises the following question: What accounts for the relatively high inter-rater agreement using the figure drawings, since adjustment ratings are not correlated with IQ, and contrary to one of Whitmyre's conclusions, we suspect that degree of artistic excellence accounts for only a small measure of the agreement?

## Summary

Three judges independently rated figure drawings and group Rorschachs of 96 freshman college students by dividing them into three equal piles according to level of personal adjustment. There was considerable inter-rater agreement, but it was relatively high for the figure drawings and much lower for the Rorschach.

Using pooled ratings, the correlation between the figure drawings and the Rorschach was fairly low but statistically stable beyond the .01 level.

The question was raised: What accounts for the relatively high inter-rater agreement using the figure drawings, since adjustment ratings do not correlate with IQ, and contrary to Whitmyre's finding, we suspect that degree of artistic excellence accounts for only a small measure of the agreement?

## References

1. Albee, G. W. and Hamlin, R. M. An investigation of the reliability and validity of judgments inferred from drawings. *J. clin. Psychol.*, 1949, 5, 389-392.
2. Whitmyre, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. *J. Consult. Psychol.*, 1953, 17, 421-424.

# THE EFFECT OF VARIOUS INTERVIEW TECHNIQUES IN EVOKING FEAR RESPONSES

CHARLES WINDLE, HOWARD MC FANN, AND JOSEPH WARD[1]

*Fort Benning, Georgia*

## INTRODUCTION

The interview method, although being the father of, and frequently the validating criterion for, the personality inventory, has remained largely free of the methodological study given its offspring. The reason for this neglect seems to be the difficulty in objectively dealing with such an unstructured procedure as the interview. The need is clear, however, for additional information about biases involved in using this method and the effectiveness of particular interview techniques. The problem of how to most efficiently evoke responses which the interviewee is prone to suppress or repress is probably the basic problem that interviews have attempted to solve. A large number of techniques have been advocated to promote truthfulness in interviewees' responses (5, 6, 10). Very little work has yet been done in experimentally evaluating these various techniques (3, 10).

The present study is concerned with the relative effectiveness of 3 interview techniques (6) for challenging doubtful statements:

(1) *Disagreement.* The interviewer disagrees with the statement made by the subject.

(2) *Amplified Agreement.* The interviewer goes beyond the statement made by the subject, overextending it.

(3) *Surprise.* The interviewer shows surprise at the statement made by the subject, thereby implying incredulity.

## BACKGROUND

The first essential in experimenting with the effectiveness of various interview techniques in evoking truthful responses is to establish a known condition in the subject. This was done in the present study by using subjects who had experienced a situation which seemed almost certain to be at least somewhat fearful. This fearful experience was jumping from a 34-foot mock tower during Airborne training.

> Airborne trainees are given practice jumps from a 34-foot mock tower, a structure in which a simulated aircraft fuselage is located 34 feet above the ground. The trainee wears a harness which is fastened by risers to a trolley device running along an overhead cable. Upon command he jumps from the tower and falls about 8 feet before receiving the simulated opening shock which arrests his fall. He then rides the cable to a sawdust mound where he is released from the trolley risers.

That this experience is fearful to all men cannot, unfortunately, be completely guaranteed. That most men admit some fear on early mock tower jumps has been demonstrated (2, 4, 8). Several investigators (1, 7, 9) have argued that such experiences are necessarily fearful since the "instinctive fear of falling" is involved. It has been assumed, then, that everyone is at least somewhat afraid when undergoing the experience of jumping from the 34-foot mock tower. However, some men may tend to classify whatever emotional reaction they experience as "excitement," a "thrill," "shyness," "nervousness" or various other terms which are usually thought to designate different emotional states.

---

[1]The research reported here was conducted by the authors while they were employed by The George Washington University, Human Resources Research Office, operating under contract to the Department of the Army. Opinions and conclusions are those of the writers and do not necessarily represent views of the University or the Department of the Army.

[2]These responses by the interviewer occasionally varied slightly in wording to be appropriate to the preceding statement of the interviewee.

## Procedure

Enlisted men who had been eliminated from the Airborne training program were interviewed by one of three interviewers shortly after their elimination. First, subjects were told that their responses would be considered confidential and would not affect their status or appear on their official records. Their names and reasons for elimination from training were then taken. Following this, they were asked the standard question. "The first time you went up in the mock tower were you scared to jump?". This rather blunt question was used to make more men deny fear initially. If the subject admitted to being "nervous," "excited" or some condition other than "scared," he was asked "Yes, but were you scared?" Only subjects who denied being scared were employed in the rest of the interview experiment. For these subjects the interview continued with one of the three experimental techniques in Table 1.

Table 1. Experimental Interview Techniques

| Technique | Response by interviewer to subject's denial of fear[a] |
|---|---|
| Disagreement | "You can't help being scared. Everybody's scared. Now, tell me, how did you really feel the first time you went up in the tower?" |
| Amplified Agreement | "You mean you felt the same way the first time in the tower that you did when making a 2-foot jump." |
| Surprise | "What? You mean you weren't scared the first time?" |

Because trainees are eliminated from training in a non-random order and because the three interviewers found themselves available in a non-random fashion, it was necessary to arbitrarily establish a procedure to randomize the application of the interviews among interviewers and among eliminated trainees. This was done by having each interviewer apply the interview techniques to trainees in an ABCA BC order. Each interviewer proceeded until 30 subjects, 10 in each group, had been seen who denied being afraid when asked the standard question.

## Results

The results of the use of the three interview techniques for challenging denials of mock tower fear are shown in Table 2. It is immediately apparent that the "disagreement" technique is superior to the other two techniques in eliciting admissions of fear.

Table 2. Number of Men Admitting or Denying Mock Tower Fear When Their Initial Denial of Fear Was Challenged in Any of Three Ways

| Response | Interview Challenging Technique | | | Total |
|---|---|---|---|---|
| | Disagreement | Agreement | Surprise | |
| Admission of Fear | 23 | 9 | 9 | 41 |
| Denial of Fear | 7 | 21 | 21 | 49 |
| Total | 30 | 30 | 30 | 90 |

$X^2 = 17.56$; $P < .001$

It is frequently found in interview studies[10] that interviewers differ in the results they obtain. The present study is no exception, as is shown in Table 3. Although the interviewers differed in evoking admissions of fear from the subjects,[3] they did not differ in finding the disagreement technique most effective.

[3]This difference may be due either to the different responses evoked by different interviewers or to different standards of classifying borderline responses.

TABLE 3. NUMBER OF MEN ADMITTING OR DENYING FEAR TO EACH OF THREE INTERVIEWERS (WI, WA, M).

| Response | Response to Standard Question | | | | Response to Probing Question | | | | Overall Response | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Wi | Wa | M | Total | Wi | Wa | M | Total | Wi | Wa | M | Total |
| Admission of Fear | 25 | 43 | 40 | 108 | 11 | 19 | 11 | 41 | 36 | 62 | 51 | 149 |
| Denial of Fear | 30 | 30 | 30 | 90 | 19 | 11 | 19 | 49 | 19 | 11 | 19 | 49 |
| Total | 55 | 73 | 70 | 198 | 30 | 30 | 30 | 90 | 55 | 73 | 70 | 198 |

$X^2 = 2.57$; $P < .30 > .20$ $X^2 = 5.73$; $P < .10 > .05$ $X^2 = 6.72$; $P < .05$

## DISCUSSION

There is little question about the validity of the finding that the disagreement interviewing technique used in this study was more effective in evoking admissions of fear responses then either the amplified agreement or surprise technique. There are, however, two questions about the meaning of these results. First, was the "successful" interview technique employed in this study really a disagreement method, or was there, in addition to disagreement, some other variable? The disagreement technique used was not a simple, direct disagreement such as: "I don't believe you," "You were, too, scared," or "The hell you weren't scared." Instead, while disagreement was only implied, a general reason for disagreeing was given and an acceptance of fearful responses by the interviewer was indicated ("You can't help being scared. Everybody's scared."). The present findings may be due largely to the promised acceptance by the interviewer, rather than to his disagreement. It should be noted, however, that acceptance of a contrary response seems to be a more natural part of the disagreement approach than either an amplified agreement or surprise approach.

Second, to what extent can this finding be generalized to other interviewing situations? Enlisted men eliminated from Airborne training are a very unique population and the admission of fear of jumping from a 34-foot mock tower is an exceedingly unusual interview topic. However, as in most interview situations, there are reasons which the subjects had for suppression and repression of mock tower fear. If a particular interview technique works best in this suppression-repression situation, it would also be the technique expected to work best in other situations involving suppression and repression.

## SUMMARY

Three interviewing techniques were employed in an attempt to evoke admissions of fear from 90 Airborne trainees who denied being afraid of jumping from the 34-foot mock tower used in training. The technique of disagreeing with the subject's denial of fear was found to be more effective than the technique of showing surprise or the technique of over-amplifying his statement.

## REFERENCES

1. BARRETT, G. Psychologic armoring for the Air Forces. *War Med.*, 1944, *5*, 142-145.
2. BASOWITZ, H., KORCHIN, S. J., AND GRINKER, R. R. Response to a life stress: The experience of anxiety. Paper presented at the APA Meeting, Cleveland, Sept., 1953.
3. BUGENTAL, J. F. T. Explicit analysis of topical concurrence in diagnostic interviewing. *J. clin. Psychol.*, 1953, *9*, 3-6.
4. FINAN, J. L. *Experimental evaluation of the procedure of giving paratroop trainees an early mock-up tower jump.* War Department, Research Branch, Information and Education Division, Washington, D. C., Report No. B-168. Oct. 1945.
5. KINSEY, A. C., POMEROY, W. B., AND MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1949.
6. OLDFIELD, R. C. *The psychology of the interview.* London: Methuen and Co., 1941.
7. REINARTZ, E. G. Some mental aspects of aviation medicine. *Aviation Med.*, 1943, *14*, 75-83.
8. WALK, R. D. Self-ratings of fear as a research instrument in fear-invoking situations. Paper presented at the EPA meetings, New York, April, 1954.
9. WEGROCKI, H. J. Anxiety and plane flight. *Psychoan. Rev.*, 1946, *33*, 1-36.
10. ZUBIN, J. *Experimental abnormal psychology.* (Preliminary mimeographed edition). New York State Psychiatric Insitute, 1953.

# A FACTOR ANALYTIC STUDY OF THE ADJUSTMENT OF CHRONIC SCHIZOPHRENICS

WILSON H. GUERTIN

*Veterans Administration Hospital**
*Knoxville, Iowa*

## Problem

Factor analysis has helped us develop concepts and classifications in the domain of schizophrenia. The conventional applications of factor analysis to a study of schizophrenia aim at syndromatic description [1, 3, 7, 9]. On the other hand, transposed factor analyses are directed at studying the intercorrelations between schizophrenics to identify the *types* of schizophrenics in the sample [4, 5, 6]. Three major types of schizophrenics are invariably encountered [4, 5, 6]. In Jenkins' descriptive scheme they are called: Withdrawn, Disorganized, and With Psychotic Reorganization [8]. These types have been revealed by an analysis of both symptom data and Bender-Gestalt performance.

Will the same three types appear when more basic features of adjustment are analyzed? Employing a unique rating scale, for example, the Hospital Adjustment Scale (HAS), developed by Ferguson, McReynolds, and Ballachey [2], might provide a different perspective for viewing the domain of schizophrenic behavior.

## Procedure

Subjects were 24 resident male patients with varied schizophrenic subtype diagnoses and one hypothetical normal based upon scores in the expected directions. All subjects were chronically ill, and no attempt was made to select them as representative of the hospital. Rather, they were chosen on the basis of the availability of the HAS's, with some attempt to include the less frequently encountered diagnostic types. However, diagnoses were not recent, and the sub-types no longer retained much descriptive value.

The scale scoring was obtained in the recommended fashion from the psychiatric charge aides [11]. Tetrachoric intercorrelations between the subjects were obtained from four-fold contingency tables and charts. The resulting intercorrelation matrix was factored by the multiple-group centroid method [12]. Communalities for those individuals within a cluster were estimated by employing the individual's largest correlation within the cluster. Those outside the clusters had communality estimates based upon the highest intercorrelation in the whole column.

## Results

No general factor corresponding to schizophrenia was encountered, but one unusually large factor did appear. Another cluster appeared almost at right angles to this unusually large cluster. Extraction of these two clusters led to the factor matrix seen in Table 1. Decimal points have been omitted.

The HAS has three subdivision scores as given in Table 2. The table shows the correlations of each person's subdivision scores and total score with each of the two factors.

The original communality estimates accounted for about 85 per cent of the total variance of the individuals. The final factoring accounted for about 80 per cent of this originally estimated communality. Thus, approximately 68 per cent of the behavior of the schizophrenics sampled by the scale could be described in terms of the two factors.

---

*Author now in Medical Service to Supreme Bench, Baltimore 2, Maryland.
Dr. Theodore T. McKnelly's contribution in collecting the original data is hereby acknowledged.

TABLE 1. FACTOR MATRIX SHOWING TYPES OF SCHIZOPHRENIC ADJUSTMENT

| Subject | Factor A Apathetic | Factor B Withdrawn |
|---|---|---|
| 14 | 96 | 29 |
| 4 | 88 | -28 |
| 15 | 83 | -22 |
| 13 | 82 | 12 |
| 10 | 79 | 13 |
| 11 | 78 | 01 |
| 3 | 77 | 24 |
| 12 | 77 | 45 |
| 23 | 72 | 31 |
| 22 | 71 | 39 |
| 21 | 63 | 43 |
| 20 | 61 | 58 |
| 17 | 56 | 15 |
| 19 | 54 | -11 |
| 8 | 44 | 03 |
| 24 | 34 | 93 |
| 2 | 13 | 92 |
| 1 | -10 | 91 |
| 18 | 22 | 90 |
| 7 | -35 | -89 |
| 25* | 30 | -87 |
| 16 | 17 | 84 |
| 5 | 54 | -71 |
| 9 | 43 | 64 |
| 6 | 52 | 53 |

*Hypothetical normal individual.

TABLE 2. CORRELATIONS BETWEEN "HAS" SCORES AND SUBJECT'S FACTOR LOADINGS

| HAS Subsection Score | Factor A Apathetic | Factor B Withdrawn |
|---|---|---|
| Communication and interpersonal relations | .38 | -.83 |
| Care of self and social responsibility | .40 | -.69 |
| Work activities, recreation | .66 | -.71 |
| Total | .54 | -.87 |

## DISCUSSION

The extremely high communality among the subjects of the study argues well for the reliability of the test instrument. It is presumed that the easily understood, fundamental nature of the items is responsible for this reliability in scoring. The high communality also supports the accepted view that chronic schizophrenics are much alike in their ward adjustment.

The largest factor, the Apathetic type, is much like Kraepelin's concept of schizophrenia, yet seven of the schizophrenics in this sample do not show appreciable loadings on it. Once again the search for a general factor that might be described as "schizophrenia" is disappointing.

Table 2 demonstrates the high relationship between low adjustment scores on the HAS and the Actively Withdrawn factor-type. It is interesting to note that none of the subdivisions of the HAS is differentially predictive for either of the two schizophrenic adjustment types. The subdivisions were subjectively determined[10], yet seem to have some empirical support[7]. That the subdivisions have pragmatic value with schizophrenics or other types of psychiatric patients must still be demonstrated.

*Apathetic Type.* This factor appears to be the larger of the two factors and is loaded .40 or greater for 15 of the 24 subjects. Characteristic of this type of schizophrenic is the HAS item, Never asks for a pass. Unlike the withdrawn type, (he) Does not resent being questioned, Does not yell at the attendant, Does not do the opposite, and Does not refuse to do his duties. While these patients are somewhat withdrawn and inactive, as compared with normals, they are much more outgoing than the Actively Withdrawn type. These patients often appear rather stupid and insensitive. Much of their behavior can be characterized as "heedless of personal needs."

*Actively Withdrawn Type.* This type of patient, while also heedless of his needs, is not unconcerned. He is actively retreating from any contact with the world. (He) Never volunteers information, Needs supervision with dressing, Never asks for anything, and Is generally interested in nothing. (He) Does not seem to enjoy being talked to and Does not always reply when he is spoken to. (He) Does not engage in any recreation and Will not help out on the ward. It is apparent that these patients are merely vegetating and will resist any efforts to interest them in anything.

Of the three types of schizophrenics proposed by Jenkins[8], the type With Psychotic Reorganization is missing in this study. However, Jenkins' Withdrawn type does correspond well to the Actively Withdrawn of this study. Furthermore, the type designated as Disorganized seems very much like the Apathetic of this study. The absence of the type With Psychotic Reorganization is probably explained by reference to two things. First, the chronic nature of the sampling would tend to limit the number of actively paranoid individuals encountered, and second, the test items provided little opportunity for the sampling of characteristically paranoid features.

It is felt that this study has provided some cross-validation for previous studies of schizophrenia and has helped illustrate the changes in factorial constellation that may occur when the domain is viewed from a different vantage point.

## Summary and Conclusions

The Hospital Adjustment Scale ratings of twenty-four diagnosed schizophrenics and one hypothetical normal were submitted to a transposed multiple-group factor analysis. Two factors were revealed: the *Apathetic Type* and the *Actively Withdrawn Type.* Agreement with other studies is discussed.

## References

1. Blair, H. W. A factor analytic study of the social behavior of schizophrenics. Unpublished doctoral dissertation, Michigan State College, 1952.
2. Ferguson, J. T., McReynolds, P. and Ballachey, E. L. *Hospital Adjustment Scale.* Palo Alto: Leland Stanford Junior University, 1953.
3. Guertin, W. H. A factor analytic study of schizophrenic symptoms. *J. consult. Psychol.*, 1952, *14*, 308-312.
4. Guertin, W. H. An inverted factor analytic study of schizophrenics. *J. consult. Psychol.*, 1952, *14*, 371-375.

5. GUERTIN, W. H., AND ZILAITIS, V. A transposed factor analysis of paranoid schizophrenics. *J. consult. Psychol.*, 1953, *17*, 455-458.
6. GUERTIN, W. H. A transposed factor analysis of schizophrenic performance on the Bender-Gestalt. *J. clin. Psychol.*, 1954, *10*, 225-228.
7. GUERTIN, W. H. A factor analysis of schizophrenic ratings on the Hospital Adjustment Scale. In Press.
8. JENKINS, R. L. The schizophrenic sequence: withdrawal, disorganization, psychotic reorganization. *Amer. J. Orthopsychiat.*, 1952, *22*, 738-748.
9. LORR, M., WITTMAN, P. AND SCHANBERGER, W. An analysis of the Elgin Prognostic Scale. *J. clin. Psychol.*, 1951, *7*, 260-263.
10. MCREYNOLDS, P. Personal Communication, 1954.
11. MCREYNOLDS, P. AND FERGUSON, J. T. *Clinical manual for the Hospital Adjustment Scale.* Palo Alto: Leland Stanford Junior University, 1953.
12. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: University of Chicago Press, 1947.

---

# HYPOHIDROSIS: A CASE STUDY

J. L. MCCARY, PH.D.
*University of Houston*

JAMES ALBERT BROWN, M. D.
*Almeda Clinic, Houston, Texas*

## PROBLEM

A 19 year old, single, white male, 6 years older than his only sibling (female), came for treatment because of hypohidrosis—diminished perspiration—an ailment he had developed 5 months previously while he was an enlisted man in the United States Navy. His mother made the appointment for him. The patient's only complaint, other than the hypohidrosis, was that he occasionally had severe headaches, which had first developed a short time after he joined the navy and which had recurred two or three times each month. Physical and neurological examinations revealed no organic causes for the disturbances.

The patient was a handsome, well-developed, ex-high school star athlete, of mesomorphic body build, 6 feet tall, 180 pounds. He was pleasant and talked readily if questioned directly. He did not look often at the therapist but kept his eyes averted most of the time. He dressed neatly, if a little ostentatiously; and, although he was of Protestant religion (non-devout), he wore a St. Christopher medal which was prominently displayed by an open-throat sport shirt. He had a tattoo on each arm. He was of bright average intelligence (Wechsler-Bellevue). Before joining the navy he was an apprentice plumber, working for his father.

## PRECIPITATING FACTORS AND SYMPTOMS

Shortly after graduating from high school, the patient and a close friend volunteered for a four-year tour in the U. S. Navy. The second night he was in boot camp he and "a group of buddies went over the hill" because they "didn't like that taking orders." He stated that he had never taken orders from anyone before, wasn't prepared for it and didn't like it. He was given 10 days restriction at Captain's Mast as punishment. Five months later he was 6 hours absent over leave from week-end liberty, receiving this time 14 days restriction at Captain's Mast. These were the only two such occurrences he experienced while in the service.

The patient first developed his headaches 5 months after enlisting in the navy on an occasion when he was "just riding around" with his parents who were visiting him at his navy base. For this first headache he could gain no relief with medication. The best treatment seemed to be a few hours of sleep, at the end of which the headaches would be gone. These headaches persisted regularly—about 2 or 3 per month—until the patient was discharged from the navy, at which time they stopped temporarily. However, there was no relief from the other symptoms. The patient reported

that he had been "nervous" before joining the navy and not in very good health (although he could not report specific symptoms), but that he settled down after entering the service.

A short time after he developed his first headache, he was assigned to a ship where he served satisfactorily until his hypohidrosis developed. He reported that he got along well with most people aboard, officers and men, and was especially fond of his boat officer. However, he had some difficulty with his chief petty officer. His dislike for the petty officer developed because the chief "expected all the sailors to get in a hurry for him." The only other exception to his good behavior was one occasion when he had a fight aboard ship with one of the men who bothered him. He could not recall what caused him to be upset. The fight itself consisted of one punch, landed by the patient, which knocked his opponent down. Other men separated the two and nothing else came of the disturbance.

Two months after he started his sea duty, the patient and two friends made application for discharge from the navy, giving family hardship as cause. He stated that his parents were deeply in debt and he felt they needed the financial help he would be able to give them if he had a civilian job. It took 8 weeks for his request to be acted on. Although his friends were both released, one on a hardship discharge, the other on a medical discharge (ulcer), the patient's request was refused. He stated that he was not bothered by this turn of events, even though he had thought that he would be upset if he failed to get the discharge. He accounted for his failure to get upset by the fact that his parents' financial condition had improved substantially during the 8 weeks. On one other occasion the patient talked with his chaplain and tried to get a transfer to a naval base near his home, but nothing ever came of this attempt.

Three weeks before his ship was to return to the United States, the patient, while removing a small pump from his boat, began to sting all over and noticed a slight rash. He also noticed that he was not perspiring. He took off his jacket and sat down for awhile and the sting and rash disappeared. Gradually, however, the condition grew worse every time he did any work or was in any place which would ordinarily cause him to perspire. Eventually he had to sit under a fan all the time and the ship's physician sent him to a shore base hospital. He developed the first symptoms in early February and was sent to the hospital March 28, almost two months later.

The patient was in the hospital for four months, primarily for observation, although he was given one "test or shot which was supposed to cause sweating within two minutes." Several dermatologists in the hospital examined him and told him to seek psychological help, adding, however, that they did not know whether or not it would help his condition. He did not see a psychiatrist or psychologist while in the hospital.

During his four months in the hospital, the patient had "several nightmares," but could recall only two. On one occasion he dreamed he and a close friend were having a fight and several people were restraining him. He woke up screaming and fighting. In the other dream he thought someone was "laying out under a sheet" and that he "was running back and forth over them in a car."

After four months in the hospital—15 months in the service altogether—the patient was discharged. He was given a 10% disability rating because of his condition. At his home in Texas, his condition became much worse because of the excessive heat, and he gained relief from his condition only while staying in an air-conditioned room, which he left only at night, early in the morning or late afternoon. Prolonged activity or excessive heat would cause him to lose consciousness inasmuch as he could not gain relief through perspiring.

## Past History

The patient was born when both parents, who are still living, were 24 years old. He was delivered with instruments but no unusual complications resulted. As a child,

the patient had few illnesses: whooping cough at age 9 months, and thereafter, only an occasional cold or slight fever, sometimes associated with tonsillitis.

The mother, who gave most of the history on the patient, is a friendly and understanding person who has "always shown a normal concern over the boy but has never been over-protective." She spent 6-8 weeks in a tuberculosis hospital when the patient was 8 years old but returned home before her doctor dismissed her because she missed her children and husband. A practical nurse, living in the home, cared for the children while the mother was in the hospital.

The father has always been a hard working man and an adequate provider. During the patient's early life the father worked as a truck driver, and was often gone from home for a week at a time; but his absence did not appear to upset the children. His only serious illness was one incident of typhoid. When the patient was 4 years of age, his father attempted to teach him to swim by tossing him in the water. He developed a fear of water from this episode and kept the fear until he joined the navy and had to learn to swim.

The patient was an exceptionally good child when young and cried very little. He was toilet trained at an early age but apparently without unusual problems. He was considered to be a smart child, if often somewhat stubborn. Living in the city and attending schools near his home, he had a good attendance record and never got into trouble with teachers or pupils. He usually made average grades in school, although he did fail one English course in junior high school.

Nothing unusual happened during the patient's youth; but, as he developed physically, he began to show some outbursts of aggression. About the time he entered junior high school he began to have a number of fights, which he usually won. He resented being "given orders" and would cooperate much more freely if asked to do something. He also resented it strongly when anyone—especially a man—put his hands on him.

He occasionally got into trouble with the law, usually for speeding or running red lights but sometimes for gang fights and gang troubles.

The patient was an outstanding football player in high school and was offered several scholarships, but refused them because he didn't want to leave his home town and wasn't ready for college. He stayed one week at one university but decided against enrolling.

He obtained most of his sexual education from other boys at school and denies any guilt or other problems in this area. He had his first intercourse at age 14 and has continued fairly regularly since that time. He denied any homosexual experiences.

After graduating from high school, he worked for his father, who owned a plumbing concern, as an apprentice plumber. He went out almost every night and continued to have minor brushes with the law because of speeding, fighting, etc. He joined the navy when just 18, partly in order to avoid more conflict and fines with the police department.

## Course of Treatment

During the first few therapeutic sessions, the patient was seen at night since it was in the middle of the summer and he could not be in the heat, his symptoms being rather severe at that time. He was seen twice a week.

The problem was discussed with the patient for several sessions; but, because of his apparent inability to gain any understanding of the etiological factors involved, and because of his continuing problem with his symptoms, it was decided to use narcosis and hypnosis as an adjunct to psychotherapy.

During the first pentothal interview he became very tense and developed a tingling and itching sensation over his entire body. This reaction disappeared after a few minutes of suggestion. He began to perspire slightly on his upper lip and between his fingers as a result both of suggestion and of discussion of his emotional feeling toward the navy and the way he was treated while in the service. At the end of the first session the patient reported that he felt as if the rash were returning.

During the second pentothal interview, 5 days after the first, the patient discussed the guilt he felt as a result of getting out of the service the way he did. He stated that after he recovered he intended to re-enlist in the navy to show them that he is a strong person. Although the patient perspired considerably more during the discussion of his problem in the second interview than he did during the first, he reported that his condition had not improved away from the clinic, despite suggestions which had been given him while under the influence of the drug.

At the beginning of the next session the patient reported that a few nights before he had put his head on his arm in order to rest and noticed that he had perspired on his forehead. He was not given pentothal during this session, but was encouraged to discuss his reactions under the drug and how these reactions were related to his problem. He reported that for the first time he had begun to notice that he developed that same rash and itching sensation, caused by heat, when he was in an emotionally upsetting situation. He also reported that his hands had been much more moist than usual and that his skin no longer "cracked open and scaled off."

The patient was in the hospital for the next month because of "intestinal trouble" and was not able to keep his appointments. At the end of the month he returned for an appointment made for him by his mother. During this session, in which no drugs were used, he reported that he had developed two severe headaches during the past week, but that they had been relieved by pills and sleep. He discussed his problem willingly but without emotion.

During the next six sessions he discussed his problems freely and, during one pentothal session, repeated several times that, when he had tried to get out of the navy and could not, he "just had to sweat it out." When, in a waking state, he was asked to relate "having to sweat it out" to his symptoms, he only laughed, being unable to accept the association. During these sessions he was given post-hypnotic suggestions that he would perspire more, and that he would gain an understanding as to why he needed to develop his symptoms, together with their relation to his problem.

At the time of his next appointment, the patient's mother phoned to say that he had to go to the Veteran's Administration hospital for a period of observation in regard to his pension. As a result of the observation and tests, his pension was withdrawn. He reported that at this time he was perspiring profusely on the hands, feet, and under his arms.

A month later the patient returned for his psychotherapeutic session, the appointment again having been made by his mother. During this pentothal session he related readily his need for rebelling against authority, stating that before the navy he had always had his way, resorting to force if necessary. When these tactics had failed him in the navy, he recognized the need for other methods of solving his problem—thus, his symptoms. During the next few sessions he was able to discuss his "insights" during a waking state. His ability to perspire continued to improve gradually, his entire body now being involved.

Just after being released from the V. A. hospital, the patient met an attractive young woman and married her about two months later. At about the same time he took and passed his examination to be a licensed plumber (he had been studying while confined to his room). He started to work regularily just after he married and has held this job successfully since that time.

During the last few sessions, the patient discussed, in both a waking state and while under pentothal, the fact that he had developed his symptoms because he could satisfy his needs in no other way. His usual method of operating had failed and he had to develop some more adequate method of solving his problems. It was pointed out that after his needs had been satisfied—that is, after he was released from the navy—he no longer should require the specific symptoms and could give them up. This method of discussing his problems was used, along with why he selected the particular symptoms he did, and why he needed such psychological symptoms at all. After three such sessions the patient stated that he was perspiring as freely as he ever had and that his headaches had not returned.

With the disappearance of his symptoms, the patient saw no need to continue with psychotherapy, although it was apparent that he did not have a thorough understanding of his overall situation. He felt that he was well and that, since he was married, he would be "able to settle down and not have any more trouble."

Six-month and one-year check-ups showed him to be clear of symptoms; but an 18-month check-up indicated that he was still having an occasional headache when he would get upset emotionally. On his own, he took up an active sport to help control his tension. Apparently he adjusted to married life in a satisfactory manner.

### Comments

It is interesting to note the number of times the patient was involved with events and items pertaining to water. He was toilet trained early, his father frightened him of water when trying to teach him to swim, he chose the navy for his branch of the service, his chief complaining symptom was that he could not perspire, he first noticed the symptoms while removing a small pump from a boat, and he selected plumbing as his profession.

It has been stated that the mother had never shown a tendency to be overprotective; however, it was she who made the original appointment and the two appointments following his hospitalization.

Before joining the navy, the patient was always able to express his hostility openly and directly. In the navy, on the other hand, he had to conceal the hostility he felt when he was not discharged as requested, although one of his friends was discharged because of an ulcer. At this time the patient was told, or fantasies he was told, that he would have to sweat out his term of enlistment. A short time later he developed a psychosomatic disturbance which gave him an excuse for a medical discharge; kept his aggressive reactions, which he knew he should not express, within himself; and at the same time allowed him to defy authority by the simple body language of not sweating out his term in the service. His dreams were obviously an expression of his desire to release his hostility and aggression, as was his strong resentment of anyone's putting his hands on him.

---

## MENTAL PATIENTS' ATTITUDES TOWARD MENTAL HOSPITALS

OMNEYA SOUELEM

*5 Midan El Cheick Youssef, Kasr El Doubara*
*Cairo, Egypt*

### Problem

The purposes of this study were to develop an attitude scale toward mental hospitals and to apply it to mental patient samples. It is obviously important to know what mental patients think of the most important treatment setting—mental hospitals—as they are the persons most involved. A survey of the literature revealed that no scale of attitudes toward mental hospitals has been developed and no study has been made on the attitudes of mental patients toward mental hospitals.

### Method

An equal-appearing interval scale was constructed following the procedures outlined by Thurstone[1, 2]. Using the ratings of 30 judges to eliminate ambiguous

---

[1]This article is based upon a thesis submitted in partial fulfillment of the M. A. degree in Psychology at the University of Oregon. A detailed description of the development of the scale and the results obtained will be found in the thesis which is filed at the University of Oregon Library. The writer especially wishes to thank the staff members of the Veterans Administration Hospital at Roseburg, Oregon, and the State Hospital at Salem, Oregon, for their generous cooperation.

and irrelevant items, the original 307 item pool was reduced to two forms of 36 items each. An 11 item sentence completion test (Part 1) was devised as a supplementary approach to the elicitation of attitudes. The attitude scale (Part II) was scored in the usual way by totaling the scale values of the statements agreed to by the subject. In order to obtain scores for the sentence completions, they were submitted to three clinical psychologists who were asked to assign numbers to each completion: 1 (favorable toward mental hospitals), 2 (neutral), or 3 (unfavorable toward mental hospitals). Inter-rater reliability coefficients ranged from .795 to .866. A copy of the entire test is given in Fig. 1. The scale values given to the left of each attitude scale item did not appear on the test presented to the patients. The first 36 items in Part II constituted Form A of the attitude scale, the remainder constituting Form B.

FIG. 1. SENTENCE COMPLETION TEST AND ATTITUDE SCALE FOR SAMPLING PATIENT'S ATTITUDES TOWARD MENTAL HOSPITALS

Name:.....

What Do You Think About Mental Hospitals?

PART I

This is a study of what people think about mental hospitals. We would greatly appreciate your cooperation here in telling what you think and feel. Below are the beginning words of several sentences. Complete the sentences in your own words. Write in several words expressing your ideas and feelings. Since there is a limited amount of time, write quickly but be sure to express your real attitudes and thoughts.

1. I believe mental hospitals are.....
2. The worst thing about mental hospitals is.....
3. Most people think mental hospitals are.....
4. Being in a mental hospital is.....
5. Mental hospitals should.....
6. Mental hospitals make me feel.....
7. The best thing about mental hospitals is.....
8. Those persons who fear being in mental hospitals are.....
9. Those persons who hate being in mental hospitals are.....
10. Those persons who are ashamed about being in mental hospitals are.....
11. Those persons who like being in mental hospitals are.....

PART II

Here is another way of saying what you think about mental hospitals. On the following pages you will find many statements people sometimes make about mental hospitals. Read each statement and show whether you agree with the statement or not. If you agree with the statement, underline the word "agree" beside the statement. If you do not agree, underline the words "do not agree" beside the statement.

| | | | | |
|---|---|---|---|---|
| (10.3) | 1. | Mental hospitals are evil and sinful. | Agree | Do not agree |
| (5.5) | 2. | Being in a mental hospital is neither good nor bad. | Agree | Do not agree |
| (7.6) | 3. | I would hate to work in a mental hospital | Agree | Do not agree |
| (9.5) | 4. | Mental hospitals are nauseating. | Agree | Do not agree |
| (2.2) | 5. | Mental hospitals perform a useful service in our society. | Agree | Do not agree |
| (8.8) | 6. | Mental hospitals bring unnecessary misery to the patient. | Agree | Do not agree |
| (6.8) | 7. | There are too many things to do in a mental hospital; a patient gets tired. | Agree | Do not agree |
| (5.1) | 8. | Mental hospitals are neither to be enjoyed nor to be feared, they are to be accepted. | Agree | Do not agree |
| (10.4) | 9. | Mental hospitals are houses of "living" death. | Agree | Do not agree |
| (0.6) | 10. | Mental hospitals are perfect in every way. | Agree | Do not agree |
| (3.3) | 11. | Recreation in mental hospitals is very entertaining. | Agree | Do not agree |
| (2.6) | 12. | Mental hospitals are improving every year. | Agree | Do not agree |
| (8.5) | 13. | A normal person could not stand being locked up in a mental hospital. | Agree | Do not agree |
| (1.0) | 14. | Mental hospitals are a great benefit to mankind. | Agree | Do not agree |
| (9.9) | 15. | We would be better off if no one had ever thought of mental hospitals. | Agree | Do not agree |
| (6.5) | 16. | Many patients won't want to be in a mental hospital against their will. | Agree | Do not agree |

| | | | | |
|---|---|---|---|---|
| (8.9) | 17. | Most of those who enter mental hospitals might as well give up hope. | Agree | Do not agree |
| (1.9) | 18. | Mental hospitals give real help in meeting social problems. | Agree | Do not agree |
| (5.5) | 19. | Mental hospitals inspire no definite likes or dislikes in me. | Agree | Do not agree |
| (4.0) | 20. | A mental hospital is a place to rest. | Agree | Do not agree |
| (7.4) | 21. | In mental hospitals they prevent a person from doing what he wants. | Agree | Do not agree |
| (4.1) | 22. | Being in a mental hospital is a sort of vacation for a person. | Agree | Do not agree |
| (2.6) | 23. | Most patients really want to be in mental hospitals. | Agree | Do not agree |
| (4.4) | 24. | I guess the mental hospital is the right place for a patient. | Agree | Do not agree |
| (9.9) | 25. | A person going into a mental hospital should expect the worst. | Agree | Do not agree |
| (4.7) | 26. | While a person is in the mental hospital he should do what they tell him to do. | Agree | Do not agree |
| (3.1) | 27. | Mental hospitals are quite nice and restful. | Agree | Do not agree |
| (6.9) | 28. | I don't like the idea of not being able to go where one wants to go, and to do what one wants to do, in a mental hospital. | Agree | Do not agree |
| (5.7) | 29. | I don't care whether mental hospitals are good or bad. | Agree | Do not agree |
| (3.3) | 30. | I think that mental illness can be helped in a mental hospital. | Agree | Do not agree |
| (3.0) | 31. | A mental hospital is probably the best place for a mentally sick person. | Agree | Do not agree |
| (1.7) | 32. | The mental hospital is a great help to the mentally sick. | Agree | Do not agree |
| (7.8) | 33. | It is better for a mentally sick person to be treated at home rather than in a mental hospital. | Agree | Do not agree |
| (7.1) | 34. | I would dislike being forced to go to movies and dances and ball games while in a mental hospital. | Agree | Do not agree |
| (1.3) | 35. | The men who started the first mental hospitals were great contributors to humanity. | Agree | Do not agree |
| (8.1) | 36. | Mental hospitals are frightening. | Agree | Do not agree |
| (0.7) | 37. | Mental hospitals are the most admirable of institutions. | Agree | Do not agree |
| (5.6) | 38. | Both the evils and benefits of mental hospitals are greatly exaggerated. | Agree | Do not agree |
| (2.9) | 39. | Being in a mental hospital does not make a person feel he is so different after all. | Agree | Do not agree |
| (7.6) | 40. | The money spent on mental hospitals could be much better spent on schools. | Agree | Do not agree |
| (9.7) | 41. | Mental hospitals are basically immoral. | Agree | Do not agree |
| (8.3) | 42. | Most mental hospitals give patients a feeling of unrest and anxiety. | Agree | Do not agree |
| (3.7) | 43. | I think the mental hospital is doing most patients some good. | Agree | Do not agree |
| (1.3) | 44. | Basically mental hospitals are a wonderful thing. | Agree | Do not agree |
| (5.4) | 45. | I don't know what to think about mental hospitals. | Agree | Do not agree |
| (0.8) | 46. | Mental hospitals are a blessing to mankind. | Agree | Do not agree |
| (4.2) | 47. | There is nothing unusual about being in a mental hospital. | Agree | Do not agree |
| (8.9) | 48. | Families of mental patients should be ashamed of sending them to the hospital. | Agree | Do not agree |
| (9.8) | 49. | Mental hospitals have not changed much since the time when they chained patients and beat them. | Agree | Do not agree |
| (4.7) | 50. | Some patients like it in a mental hospital—Three meals a day, no worry. | Agree | Do not agree |
| (7.3) | 51. | You'd think there would be more practical ways of handling patients than mental hospitals. | Agree | Do not agree |
| (7.7) | 52. | The hospital people don't pay enough attention to individuals and how different everyone is from everyone else. | Agree | Do not agree |
| (10.4) | 53. | Mental hospitals are as bad as concentration camps. | Agree | Do not agree |
| (6.6) | 54. | If they would just let patients do what they want to do in the mental hospitals, they would all get to feeling much better. | Agree | Do not agree |
| (1.9) | 55. | A patient is definitely being helped in a mental hospital. | Agree | Do not agree |
| (8.8) | 56. | Anyone who goes to a mental hospital should be ashamed. | Agree | Do not agree |
| (2.2) | 57. | The number of patients cured in mental hospitals is rising rapidly. | Agree | Do not agree |
| (1.4) | 58. | Mental hospitals are concerned with the welfare of every patient. | Agree | Do not agree |
| (9.9) | 59. | Mental hospitals are a disgrace. | Agree | Do not agree |
| (8.2) | 60. | Mental hospitals are so poorly planned that patients hate them. | Agree | Do not agree |
| (10.3) | 61. | Mental hospitals are snake pits. | Agree | Do not agree |
| (6.8) | 62. | In the mental hospitals they try to get the patients to talk about themselves too much. | Agree | Do not agree |
| (3.2) | 63. | Although all patients do not improve in mental hospitals, most of them are helped. | Agree | Do not agree |

| | | | | |
|---|---|---|---|---|
| (2.3) | 64. | A mental hospital is a place where the patient is relieved and comforted. | Agree | Do not agree |
| (5.5) | 65. | Mental hospitals are neither good nor bad. | Agree | Do not agree |
| (4.3) | 66. | Mental hospitals are alright, after all. | Agree | Do not agree |
| (3.5) | 67. | Improvements are being made in mental hospitals. | Agree | Do not agree |
| (5.6) | 68. | One hears so many different ideas about mental hospitals that it is hard to decide whether they are good or bad. | Agree | Do not agree |
| (6.9) | 69. | There is too much time in mental hospital with nothing to do. | Agree | Do not agree |
| (7.3) | 70. | In the mental hospital, patients don't get any chance to use their own abilities to the best advantage. | Agree | Do not agree |
| (3.3) | 71. | There is a growing need for mental hospitals. | Agree | Do not axree |
| (2.9) | 72. | In mental hospitals they have very interesting things for the patients to do. | Agree | Do not agree |

## Samples

Two samples of male mental patients were used in the present study: 95 men from four wards at the Roseburg (Oregon) Veterans Administration Hospital, and 103 men from four wards at the Salem (Oregon) State Hospital. The VA sample was given the sentence completion test and both forms of the attitude scale. The State hospital sample was given only Form A. An attempt was made to obtain a representative sample of all capable and cooperative patients on the wards tested. The wards used in both hospitals included a variety of cases from newly admitted patients to old chronic cases.

## Results

The attitude scale was found to have a high equivalent-form reliability (.88 between Forms A and B). Some evidence for the validity of the scale was inferred from the correlation coefficients of .59 and .60 between averages of psychologists' ratings of the sentence completions and the two forms of the attitude scale. On the attitude scale the majority of patients in both institutions scored in the favorable end of the scale, approximately 85 per cent being above the midpoint of the scale. However, on the sentence completions according to the psychologists' ratings a smaller majority (66 per cent) fell in the favorable half of the distribution. Analysis of variance revealed no significant differences among attitude scores of patients in the various diagnostic categories. The correlation of the attitude scale scores with length of hospitalization at both institutions was low and insignificant. Likewise there was no significant relationship between attitudes and patients' ages. No significant differences in attitude were found in comparisons between the total samples at Salem and Roseburg or between comparable wards in the two mental institutions. Significant differences were found between wards within the hospitals however. In both cases the admissions wards and more active convalescent wards had more favorable attitudes toward mental hospitals than the more chronic and semi-convalescent wards.

Table 1. Mean, Standard Deviation and Range of Attitude Scale Scores by Wards in the V. A. Hospital

| | Ward 1 Admissions | Ward 4 Disturbed, Acute & Chronic | Ward 6 Open ward Convalescents & Chronics | Ward 8 Closed ward Chronics & Semi-convalescent | Total Sample |
|---|---|---|---|---|---|
| N | 14 | 17 | 32 | 32 | 95 |
| Mean | 4.2 | 4.2 | 4.4 | 5.1 | 4.57 |
| S.D. | .80 | .62 | .79 | 1.38 | 1.06 |
| Range | 3.5–6.6 | 3.3–5.3 | 3.0–6.6 | 3.3–5.3 | 3.0–8.2 |

TABLE 2. MEAN, STANDARD DEVIATION AND RANGE OF ATTITUDE SCALE SCORES BY WARDS IN THE SALEM HOSPITAL

| | Ward J Admissions | Ward M Active Treatment | Ward 5 Diabetes Cardiac | Ward 42 & 43 Closed Ward Chronics & Semi-convalescent | Total Sample |
|---|---|---|---|---|---|
| N | 31 | 15 | 18 | 39 | 103 |
| Mean | 4.0 | 4.1 | 4.6 | 4.9 | 4.46 |
| S.D. | .71 | .47 | 1.16 | 1.05 | .99 |
| Range | 3.1–6.1 | 3.4–4.8 | 3.3–7.1 | 3.2–7.1 | 3.0 7.1 |

## DISCUSSION AND CONCLUSIONS

The significant differences between wards, interpreted in the light of absence of relationship to diagnosis and chronicity in general suggests that further investigation of attitudes be directed toward some other individual patient variables (e.g. kind of treatment in hospital, educational background) and to the ward situations themselves (e.g. the kind of inter-patient relationships, staff-patient relationships, and general morale of the ward.) The greater expression of unfavorable attitudes in the freer situation provided by the sentence completion is an interesting finding. Some distortion of results toward the favorable end might have occurred because the scale was not given anonymously. Future studies might well include comparisons with a normal sample since the items were phrased so that they could be used either in a patient or non patient situation. With further refinement and application an attitude scale such as this might help hospitals assess general patient morale and individual patient response to the living and treatment situation.

## REFERENCES

1. FERGUSON, LEONARD W. *Personality Measurement.* New York: McGraw-Hill Publications in Psychology. 1952.
2. THURSTONE, L. L. AND CHAVE. *The Measurement of Attitude.* University of Chicago Press, Chicago, Illinois: 1929.

---

# THE BENDER-GESTALT RECALL AS AN INDEX OF INTELLECTUAL FUNCTIONING

ROLAND M. PEEK AND GORDON W. OLSON
*Hastings (Minn.) State Hospital*

It has been noted by some clinicians that there are quantitative and qualitative aspects of performance on the Bender-Gestalt (B-G) test which appear to be related to intellectual factors. Considering only the quantitative factors, Aaronson, Nelson and Holt[1] studied the relationship between the number of figures recalled and the mental age and intellectual efficiency as measured by the Shipley-Hartford (S-H) Scale. The results indicated essentially no relationship between the recall and the indices mentioned; the product moment correlation between CQ and number of figures recalled was in fact negative. The present study was undertaken to more completely understand the relationship between these variables.

## Procedure

From the files of the Hastings State Hospital were drawn all patient records having both a scorable S-H and B-G. Patients producing a S-H vocabulary score of less than 16 were excluded inasmuch as the CQs were not conveniently computable from the table in the manual. The S-H was administered and scored as described in the manual for that test[2] with the exception that the patient was not restricted to the ten-minute time limit. The B-G recall was obtained by requesting the patient to "draw as many (figures) as you can remember", after removal of the stimulus cards and the patient's original protocol and without the subject's foreknowledge of the recall requirement[3].

Whichever S-H subtest score (Vocabulary or Abstraction score) was higher was used as the best estimate of intellectual ability; the B-G recall score was obtained by assigning a score of one for each figure completely recalled and one-half where only a clearly recognizable element[3] was reproduced.

## Subjects

The resulting sample was composed of 77 female and 116 male neuropsychiatric in-patients ranging in age from 14 to 72 years, with a mean age of 34.67, a median age of 32.17, and a standard deviation of 12.85 years. The skewness of the age distribution was +.19.

## Results

The mean number of B-G figures recalled for the entire group was 5.09 with a standard deviation of 2.08; the mean CQ was 83.06 with a standard deviation of 16.30. The distribution of S-H raw scores for the group had a mean of 28.81 with a standard deviation of 5.67. The product moment correlation between number of figures recalled and CQ was .34, significant at a .001 level, while the correlation between the number of figures recalled and S-H raw score was .19, significant at the .01 level. These correlations indicate that in our sample the B-G recall varies more consistently and to a greater degree with intellectual efficiency as measured by the CQ than it does with intellectual endowment as estimated by the higher S-H sub-test score.

In order to determine any differences at various intellectual levels, the sample was divided into three groups of low, average, and high intelligence according to their performance on the S-H. The mean number of figures recalled was computed for each group and the significance of the difference between the means determined. The mean number recalled in the low scoring group was 4.53 while for the high group it was 5.40; the difference of .87 figures is significant at the .05 level. It was noted that 32% of the high scorers had CQs greater than 100 as compared with only 6.8% of the low scorers. Therefore, to control CQ as a variable affecting the relationship between the S-H raw score and B-G recall, only those records with CQs of less than 100 ($N = 73$) were sorted into high and low scoring groups. When this was done the low scorers ($N = 41$) had a mean recall of 4.34 while the high scorers ($N = 32$) had a mean recall of 5.03; the difference of .69 figures is not statistically significant. Considering only those individuals with CQs of 100 or more ($N = 35$) the correlation between S-H raw score and B-G recall drops to −.035; this group produces a mean recall of 6.19 figures, which is higher than for any of the other groups examined.

Comparison of these findings with those of Aaronson, Nelson and Holt[1] shows that although their age range is more restricted, the median ages of both samples are almost identical, and they apparently performed very similarly as a group on the S-H. However, for their 50 males they found no significant correlations between the two measures, reporting a product moment correlation (not significant) of −.10 between number of figures recalled and CQ, and a correlation of .03 between number of figures recalled and S-H raw score. The difference in findings may reflect the greater number of subjects in the present study, or the somewhat greater heterogeneity of population (wider age range, both sexes, in-patients rather than out-patients, and

use of a slightly lower vocabulary score criterion for inclusion in the study). The difference does not seem to be a result of divergent proportions of subjects achieving CQs of 100 or more since the S-H behavior of the subjects in both groups is apparently very similar.

### Personality and History Factors

Since a consistent relationship had been established between B-G recall and S-H CQ, it was felt that it might be profitable to examine those cases where there was a marked discrepancy between the two measures. Accordingly, two groups of these deviant cases were formed: (1) those cases with a CQ of 100 or more but a recall of 4 or less figures, and (2) those with a recall of 6 or more figures and a CQ of 75 or less (which Shipley reports as occurring in less than 5% of his "normals"). The histories of the individuals in these groups were inspected for the presence or absence of the following items: length of hospitalization; number of previous hospitalizations for mental illness; committed or voluntary status; involvement with the law; physical assaultiveness; verbal assaultiveness, suicidal attempts; suicidal threats; sex deviation; marital status; and diagnosis. In addition to the history items, the MMPI T-scores for these cases were tabulated and a mean profile drawn for each group. Although the N's were considered to be inadequate for definitive statistical analysis (N = 9 and 16, respectively), it was felt that these results might provide some indications of trends which will be more extensively investigated as more cases accumulate.

The following marked trends were noted: The high CQ-low recall subjects tended to have a longer period of hospitalization, but with fewer prior hospitalizations; fewer of them were voluntary patients, and there was less involvement with the law. They had a greater tendency to threaten suicide. There was no overt sex deviation in this group, though this was present in 37.5% of the others; on the other hand, fewer of the high CQ-low recall subjects had ever been married. On the MMPI, this group tended to be less defensive as measured by the L and K scales, particularly L, and to show more anxiety and internalization of their emotions as measured by the Welsh AI and IR indices[4]. They admitted to more deviation on the clinical scales, most markedly on D, Pa, Sc and Si. They scored less than the other group on none of the clinical scales although the mean T-scores were approximately equal on Hs, Hy, Pt, and Ma. The mean Mf score for this group was considerably higher than for the other males, though no difference between the means was shown by the females of the two groups. Each of the two samples contained about half men and half women, which is a somewhat greater proportion of females than in the total group of 193 originally studied.

It appears from these data that the high CQ-low recall group is more overtly mentally ill and more willing to admit to such. They perhaps acquire less incidental learning but function more effectively in the area of abstract reasoning and judgment; they appear to react more ideationally and to introject more, in contrast to the more behavioral reactions of the other group. These tentative findings strongly suggest a complex interrelationship between the variables affecting performance on the B-G recall and on the S-H, and which are related to both intellectual and emotional factors.

### Summary

A study of the relationship between intellectual function on the Shipley-Hartford scale and the Bender-Gestalt recall for 193 neuropsychiatric in-patients reveals a correlation of .34 ($p < .001$) between intellectual efficiency as measured by the S-H CQ and the number of B-G figures correctly recalled; a correlation of .19 ($p < .01$) was found between the S-H subtest having the higher raw score and the B-G recall. The difference in number of B-G figures recalled for high and low scorers on the S-H subtest is .87 figures ($p < .05$). Patients with higher intelligence tended to get more CQ's above 100; when CQ is controlled the difference between number of figures recalled in high and low-scoring groups is statistically insignificant, as is the resulting

correlation between intellectual level and number of figures recalled. It therefore appears that the B-G recall is quantitatively associated with intellectual efficiency but not necessarily with intellectual ability.

A preliminary survey of the few exceptions to the above general findings (those cases with high CQ-low recall and those with low CQ-high recall) reveals a number of personal history items and personality factors (as evidenced on the MMPI) which appear to differentiate between the two groups.

### References

1. Aaronson, B. S., Nelson, S. E., and Holt, S. On a relation between Bender-Gestalt recall and Shipley-Hartford scores. *J. clin. Psychol.*, 1953, 9, 88.
2. Manual of Directions and Scoring Key, Shipley-Institute of Living Scale for measuring intellectual impairment. Hartford, Conn.: The Institute of Living, 1946.
3. Peek, R. M., and Quast, W. *A Scoring System for the Bender-Gestalt Test.* Hastings, Minn. (Box 292) and Minneapolis, Minn. (2810 42nd St. W.): Authors, 1951.
4. Welsh, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, 16, 65-72.

---

## EMPATHY OR PROJECTION?

RICHARD C. COWDEN

*Veterans Administration Hospital, Gulfport, Miss.*

### Problem

An attempt has been made recently by Dymond to measure empathy by means of the TAT[1] and later by developing an empathy scale[2]. She states[3], "Projection seems to be an antithetical process to empathy since projection involves the attribution of one's own wishes, attitudes, and behavior to something or someone other than the self." This study considers the following problems:

1. To what extent is one person, familiar with a second person, able to predict how that second person will answer the questions on the California Test of Personality? (Empathy)
2. To what degree is this prediction more projective than empathic? That is, is the individual's prediction of another related more to his attitude toward himself than toward the other person he is rating?
3. Is there any difference as to sex in the ability to empathize?

Empathy is defined as temporary identification with a person for the purpose of anticipating what that person is going to do in certain situations. In terms of the present problem, it is the ability of one person to predict how his or her marital partner will answer the questions on a personality inventory. By projection is meant attributing one's own qualities to another. In terms of the present problem, it is the degree to which a person's prediction of his mate's responses on a personality inventory corresponds more to his own responses than to his mate's own responses.

### Procedure

The method consisted in having twenty-nine married couples fill out twice the California Test of Personality, Form A, for adults. The subjects were married for at least one year and had at least one year of college. None of the subjects were over thirty-five years of age. They were not asked to give any identifying information but merely to specify who was answering the questions and for whom they were being answered. They were given an addressed, stamped envelope with which to mail the completed inventories to the author. The material received consisted of the following:

A. Husband answers own inventory.
B. Husband answers inventory as he thinks his wife did.
C. Wife answers own inventory.
D. Wife answers inventory as she thinks her husband did.

To determine the degree of empathy that existed, B was compared with C and A was compared with D. The inventory was analysed in terms of three categories including the total score and two sub-categories of self-adjustment and social adjustment.

To determine the degree to which the rating of the mate was more projective than empathic the correlation on each of the three categories between A and B was compared with B and C and the correlation between C and D was compared with A and D.

To determine the sex difference in ability to empathize, the empathic ability of the women was compared with that of the men. A and C, the self ratings, were correlated to determine the degree of similarity that existed between the married couples. The inventories were then randomly matched to establish a form of control in the study. The empathy and projection correlations on the randomly matched inventories were compared to determine if the results obtained in this study were more than a result of chance.

## Results and Discussion

Product-moment correlations were computed to determine the relationship existing between an individual's self rating and the corresponding rating of the individual by the mate. Since the correlations on the projective measures in this study were usually as high, or higher, than the empathy correlations it was decided to compute partial correlations to cancel out the effects of projection upon empathy; or vice versa. The results are presented in Tables 1 and 2.

Table 1. Projection and Empathy Correlations on Total Score Category

| Group | Empathy | | | | Projection | | | |
|---|---|---|---|---|---|---|---|---|
| | $r$ | P | $r_{12 \cdot 3}$* | P | $r$ | P | $r_{23 \cdot 1}$* | P |
| Females | .56** | .01 | .37 | .10 | .55** | .01 | .35 | .10 |
| Males | .70† | .01 | .52 | .01 | .76† | .01 | .63 | .01 |

*Projection partialed out.
**Correlation between A & D.
†Correlation between B & C.

*Empathy partialed out.
**Correlation between A & B.
†Correlation between C & D.

Table 2. Correlation between Husband's and Wife's Self-rating (A and C) on all Categories

| Category | $r$ | P |
|---|---|---|
| Self-Adjustment | .44 | .05 |
| Social Adjustment | .65 | .01 |
| Total Score | .55 | .01 |

The argument might be raised that the items on the inventory are such that they would be answered correctly whether or not the ratee was known. However, if this were so then a random matching would be expected to produce results which would approach those obtained from the married couples. All of these correlations were negative. This suggests that the significant results obtained are not simply a function of the inventory but rather an actual measurement of one person's knowledge of another.

The most interesting element to come from this study is the possibility that empathy and projection, as defined in this study, are not necessarily contradictory. In this study, they appear to exist side by side and are coexistent. Projection is not an antithetical process to empathy as Dymond believes. The high degree of similarity between self ratings (Table 2) helps to explain how empathy and projection can exist side by side. For example, if an individual actually were able to predict accurately how his mate would answer the inventory, this might well co-exist with high projection as measured here; e.g., relationship between an individual's self-rating and his rating for the mate. That is, the degree of similarity of responses between the couple would show up as projection even though empathy was actually involved. A comparison of Tables 1 and 2 reveals this. Thus this similarity would appear as projection because of the method used in measuring it in this experiment.

Another result of interest is the differences obtained between the scores on self-adjustment and on social adjustment. The females showed relatively less insight into their husband's actions or attitudes in the realm of social adjustment (P, .25) and their greatest projection occurred in this area (P, .02). This might be explained by the fact that the inventory is loaded with questions concerning work and community relations. It is in this area that the females might be expected to show the least insight into their husband's attitudes and feelings. That is, they would be less able to identify themselves temporarily with their husbands in this area. It may be that the husband has greater opportunity to observe his wife, than she to observe him, in this area.

Notcutt and Silva[4] suggest that the greatest success in predictability should be in those areas in which the greatest similarity of behavior or attitudes exists. The results for females in the present study do not confirm this; the results obtained on the males do support this. If one recalls that the males do much more projecting (P .01) than the females (P .10), the results of Notcutt and Silva can be explained in terms of projection rather than insight or success in prediction.

## Summary

This study consisted of having twenty-nine married couples answer a personality inventory for themselves and also one in which they attempted to predict how their marital partner would answer the questions. The degree of accuracy of prediction was considered a measure of empathy. Also, the amount of projection was determined. In general, the males were far superior to the females in ability to predict and showed greater insight than the females. The males also did more projection. These results suggest that empathy and projection are not necessarily contradictory and tend to co-exist.

## References

1. Dymond, R. F. A preliminary investigation of the relation of insight and empathy. *J. consult. Psychol.*, 1948, 12, 228-233.
2. Dymond, R. F. A scale for the measurement of empathic ability. *J. consult. Psychol.*, 1949, 13, 127-133.
3. Dymond, R. F. Personality and empathy. *J. consult. Psychol.*, 1950, 14, 343-351.
4. Notcutt, B. and Silva, A. L. M. Knowledge of other people. *J. abnorm. soc. Psychol.*, 1951, 46, 30-38.

# SCORING COLORED RESPONSES IN PAINTINGS[1]

JOHN A. POPPLESTONE

*Washington University, St. Louis, Mo.*

## Problem

Frequently interpretations of paintings are perceptive and acute but their "intuitive" bases have not been developed into a communicable scoring system. The present study attempts to contribute information toward such a system by comparing the use of color in Rorschach responses and free choice paintings. Since this study is mainly methodological, the relatively easily identified and interrelated variables of color and psychosis were selected. If a close relationship should be found in the use of color in the two situations, the significance of color in paintings could be inferred from its Rorschach concomitants. If such a parallel is not discovered, the search for the determiners of painting responses must be sought in other areas.

## Method

The subjects were 20 residents of a public institution for psychotics, selected consecutively from newly admitted patients. The only basis for exclusion was known cerebral pathology and mental deficiency. The chronological age range was from 16 to 65 years with a mean of 40 years. There were ten men and ten women, all native born monolinguals.

The Rorschach test was administered and scored according to Klopfer[(1)]. In order to secure the paintings each patient was provided with white paper, 12″ x 18″, a brush, and a tray of tempera paints containing randomly placed black, white, red, blue, yellow, and green pigments. The instructions, which emphasized free choice of subject matter and encouraged productivity, were individually adapted to the comprehension of each patient. No time limit was imposed but one subject was interrupted when it became apparent that he would have painted indefinitely.

The author noted the choice of colors and the time interval during which each choice was utilized, all verbalizations, and any atypical behavior. Time and space limitations prevented an accurate counterbalancing of the Rorschach and painting sessions but nine subjects received the Rorschach first and eleven painted first.

In order to compare the paintings and Rorschach, an adaptation of Rorschach criteria for scoring color responses was evolved and applied to the paintings. Although interest in this study was focused on color it was found essential to include the use of achromatic color in order to make certain comparisons. In painting, bright color is often appropriately supplemented by white, gray, and black to indicate outlines, shading, or the realistic representation of achromatic forms such as trunks and branches of trees. Therefore C′ scores were incorporated in some measures of the painting responses. Following are the various scoring determinants and the criteria which were used to identify them:

| | |
|---|---|
| FC & FC′ | Pigment in a definitely recognizable form and in agreement with the subject's color and form descriptions. |
| FC– & FC′– | Pigment used to make a form unrecognizable as compatible with the verbal elaboration of a definite form. |
| CF & C′F | Pigment used to make an indefinite form compatible with the subject's verbal interpretations. |
| C & C′ | Pigment used without any definite form consideration either in the painting or through verbal reference. |

[1]This research was conducted at The Mental Health Institute, Independence, Iowa. The author is now a graduate student at Washington University, St. Louis, Missouri. The late Max E. Witte, M. D., Superintendent of The Mental Health Institute at the time of the study, originally suggested this investigation and made it possible by his cooperation in freeing staff time.

## Results

A comparison was made between intellectual control on the Rorschach and on the paintings. This was expressed as the per cent of adequately structured responses (FC and FC') of the total number of colored responses. The skewness of both the Rorschach and painting percentages was so marked that the customary statistical measures were inapplicable. Comparison of the responses of individual subjects showed such disparate achievements that performance on one task cannot be used to predict the other. For example, seven subjects received 0% on the Rorschach and 100% on the paintings. Only three individuals obtained equal percentages on both tasks.

In order to describe more completely the use of color, the amount of interaction with color was computed. On the Rorschach, the customary method of quantification, the per cent of colored responses, was employed. On the paintings the percentage of time spent on the application of chromatic paints was determined. The relationship between these two percentages was quantified by the rank difference method. Since this rho was found to be −.12, it would appear that interaction with bright color is not related in the two situations.

This failure to find a relationship may be a methodological artifact, but measurements in both situations appear comparable in that both describe the time during which attention was given to color. Ruesch and Finesinger[2] found more positive results in a research that has the same aims as this paper, but which differed methodologically in two important ways. First, they estimated the size of the colored paintings even though they were irregular areas. Secondly, topics for art products were dictated. The patient-structured Rorschach responses were thus compared with experimenter-structured drawings. In addition, the fluid appearance of the ink blots was compared to the drawings made with a sharp point. The discrepancy between the results obtained by Ruesch and Finesinger and the present study may well be the result of these methodological peculiarities. The present technique appears to be a more valid method of quantification in that it measures accurately one aspect of a response, its duration, as opposed to a method that cannot measure completely its aspect of interest, irregular size. Secondly, since the amount of color interaction was measured as a function of time, it more closely approximates the Rorschach technique of dealing with the frequency of color responses rather than the size of the area producing the response.

Although the quantitative analysis failed to reveal similar response tendencies on color responses in painting and the Rorschach, a qualitative study revealed three different groups, each homogeneous in its use of color but deviant from the larger group. All Rorschach records were examined to see if any behavior on that test was similarly homogeneous and confined to the subjects in the three subgroups. This analysis indicated that Rorschach rigidity is apt to be expressed in mono-chromatic paintings but is not confined to this technique, that the indiscriminate use of all available colors is accompanied on occasion by Rorschach bizarreness, and that there is a parallel between constructive color mixing and Rorschach adequacy. These three groupings of subjects with idiosyncratic color mannerisms thus indicated a more intimate relationship between Rorschach and paintings than the gross comparisons. This technique has the disadvantage of reducing the authority of the already small $n$ but does suggest some communality of behavior in the two situations. Further, it identifies variables that can be utilized in further research into the development of a communicable description of factors which contribute to color use in painting.

## Summary

Because the scoring of paintings has been principally "intuitive", an attempt was made to devise a scoring system for color use in paintings. This was applied to the paintings of 20 mental patients and the results were compared with their color responses to the Rorschach. Quantitative measures of ability to structure color and

the responsiveness to it on the two tasks failed to reveal a relationship. However, qualitative comparisons indicated that Rorschach rigidity, adequacy, and bizarreness tend to be expressed in color use mannerisms in painting.

### References

1. Klopfer, Bruno and Kelley, Douglas. *The Rorschach technique.* Yonkers-on-Hudson, New York: World Book, 1942.
2. Ruesch, Jurgen and Finesinger, Jacob. The relation of the Rorschach color responses to the use of color in drawings. *Psychosom. Med.*, 1941, 3, 370-388.

---

## RORSCHACH AND LEVY MOVEMENT RESPONSES: A RESEARCH NOTE

GERALD F. KING

*Veterans Administration Hospital*
*Battle Creek, Michigan*

### Problem

One result of the wide interest in the Rorschach human movement response (*M*) has been the attempt to devise techniques which allow the study of this response independent of other Rorschach variables (e.g., chromatic stimulus variables). The Levy Movement Blots[6, 7, 8] represent one such attempt to provide a more controlled analysis of *M*[1]. This instrument departs from the Rorschach procedure in two principal ways: (a) the blot stimuli are all achromatic finger paintings, and (b) the instructions provide the subject with a definite set to perceive human movement. Little research has been reported for the Levy Movement Blots, and the basic question arises as to whether this method actually measures the same psychological function as that represented by the Rorschach *M*. The data presented in this article are addressed to this question.

### Method

The test results were obtained from another study[5] in which the Rorschach, Wechsler-Bellevue Verbal Scale, and the Levy Movement Blots were administered, in that order, to 100 male neuro-psychiatric patients shortly after their admission to a VA neuro-psychiatric hospital. The instructions were in accordance with Beck[4] for the Rorschach, and Zubin[8] for the Levy Movement Blots. The patients were individually screened prior to the testing to insure that the sample would consist of subjects who were "non-organic", cooperative, minimally confused, average or above in intelligence, and not over 45 years old.

On the basis of Rorschach performance, High-*M* and Low-*M* groups, consisting of 30 subjects each, were formed from the sample. The criteria were 3 or more *M* for the High-*M* and 0 or 1 *M* for the Low-*M* group. The two groups were equated for age, education, IQ, degrees of confusion and cooperation, diagnostic status, and number of Rorschach responses.

The previously proposed scoring systems for the Levy Movement Blots[2, 8] present certain difficulties for a direct comparison of *M* obtained from this method with that elicited by the Rorschach. The following two Levy measures were devised for comparative purposes: number of rejections and mean reaction time. A rejection was defined as the failure to produce a response involving *human* movement on any card. Since the Levy Movement Blots consist of 7 cards, with one being pre-

[1]Another technique for investigating *M* is the *M*-threshold method developed by Barron[3], which utilizes a series of ink blots gradated for stimulus threshold.

sented in three positions and one being repeated, the possible score for a subject ranged from 0 to 10 rejections. The mean reaction time was computed by summing the reaction times (in seconds) for each subject and dividing by the number of cards. This measure was based only on cards that elicited *M*. The assumptions underlying the comparisons were as follows:

1. The High-*M* group should have fewer rejections on the Levy Movement Blots than the Low-*M* group; and
2. The High-*M* group should have faster mean reaction times on the Levy Movement Blots than the Low-*M* group.

## Results and Discussion

The distributions of the Levy measures were dichotomized at or near the medians, and fourfold tables were utilized. An examination of Tables 1 and 2 suggests a trend for fewer rejections and lower mean reaction times to be associated with the High-*M* group, which is consistent with the rationale of the analysis. However, computed chi squares reveal that neither trend is statistically significant. Thus, it can be said that, with the sample and measurements used, no evidence was found for the existence of a relationship between the perception of *M* on the Rorschach and perception of *M* on the Levy Movement Blots.

Table 1. Comparison of the High-*M* and Low-*M* Groups in Terms of Rejections on the Levy Movement Blots

| Groups | Rejections | |
|---|---|---|
| | None | 1 or More Cards |
| High-M | 21 | 9 |
| Low-M | 16 | 14 |

$x^2 = 1.76$ (N. S.)

Table 2. Comparison of the High-*M* and Low-*M* Groups in Terms of Mean Reaction Times to the Levy Movement Blots

| Groups | Mean Reaction Times | |
|---|---|---|
| | Above Median | Below Median |
| High-M | 13 | 17 |
| Low-M | 17 | 13 |

$x^2 = 1.07$ (N. S.)

Another study by Allen, Ray, and Poole [2] has reported no relationship between the number of Rorschach *M* and the number of Levy *M* with a group of college students. Although the *number of M* seems to be a Levy measure similar in principle to the *number of rejections*, the present data were further analyzed using this Levy measure. It was found that there was also no difference between the High-*M* and Low-*M* groups in terms of the number of Levy *M*.

Attention is focused upon two areas, the blot-stimuli and the test instructions, in accounting for these negative results. Besides possible differences in ambiguity existing between the two series of stimuli, the Levy Movement Blots (unlike the Rorschach) have only blots that are achromatic in nature. In regard to the latter variable, some evidence is available to indicate that color has no appreciable affect upon the production of *M* [1]. There is also the matter of contrasting instructions or sets. On the Rorschach the subject in essence is asked, "Tell me what you see", while the Levy Movement Blots provide a much more restricted set as attested by these instructions: "What are the people doing?" It is suggested that this factor, differential sets, is the crucial one in producing the discrepant measurements of *M*.

## Summary

Data were presented in regard to whether or not the Rorschach and Levy Movement Blots measure the same psychological function. No relationship was found between the perception of *M* on the Rorschach and Levy Movement Blots. The difference in the instructions of the two tests was emphasized as the contributing factor.


## References


1. Allen, R. M. The M determinant and color in Rorschach's test. *J. clin. Psychol.*, 1953, 9, 198-199.
2. Allen, R. M., Ray, C. D., and Poole, R. C. The Levy Movement test: suggestions for scoring and relationship to Rorschach movement responses. *J. consult. Psychol.*, 1953, 17, 195-198.
3. Barron, F. Threshold for perception of human movement in inkblots, and its personality correlates. *Amer. Psychol.*, 1953, 8, 317-318. (Abstract)
4. Beck, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1946.
5. King, G. F. A theoretical and experimental consideration of the Rorschach movement response: its relation to the neuropsychiatric patient's orientation to his problem. Unpublished doctor's dissertation, Mich. State Coll., 1954.
6. Rust, R. M. Some correlates of the movement response. *J. Pers.*, 1948, 16, 369-401.
7. Rust, R. M. The Levy Movement Cards: EPA Round Table. *J. Pers.*, 1948, 17, 153-156.
8. Zubin, J. *Quantitative techniques and methods in abnormal psychology.* New York: Columbia Univer. Bookstore, 1950.


---

# AN AID IN CALCULATING Q-SORT FACTOR-ARRAYS


JAMES W. CREASER

*University of Illinois*


In Q-technique, when several person-arrays have been correlated and factored, it is often desirable to obtain a factor-array, i.e. an arrangement of items or statements which best typifies the factor.[1] These items or statements are arranged in order of their contribution to the factor loading, and from these we get an idea of the nature of the factor. Also, variance analysis is done from the factor-array. This factor-array will, naturally, be most like the person-array which has the greatest loading on the factor. It will be less like the one with the next highest loading and so on for as many persons as are to be included in the calculations.

In estimating the amount of contribution to the factor-array, the following formula, given by Spearman (1), is used to determine the weighting of each person-array:

$$w_a = \frac{f_a}{(1 - f^2_a)}$$

where $f_a$ is the factor loading for person 'a'.

Each item or statement in a person-array has a value depending upon the pile into which it was sorted. When the weighting for a person has been found by the above formula, each item-value in the array is multiplied by the weight. Then the total value of that item for the factor-array is the sum, across person-arrays, of all the weighted scores. The final factor-array is, then, the arrangement of the items in ascending order of their total scores.

Since it was felt by this author that more and more use will be made of Q-technique in psychological research, it appeared that a table from which weighting for

[1] An array, as used here, means an arrangement of items or statements in a fairly normal distribution in accordance with some set of directions.

any factor loading could quickly be found would be useful. These values are given in Table 1.

TABLE 1. VALUES FOR FINDING FACTOR WEIGHT FROM THE FACTOR LOADING.

| F. L. | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| .0 | .000 | .010 | .020 | .030 | .040 | .050 | .060 | .070 | .081 | .091 |
| .1 | .101 | .111 | .122 | .132 | .143 | .153 | .164 | .175 | .186 | .197 |
| .2 | .208 | .220 | .232 | .243 | .255 | .267 | .279 | .291 | .304 | .317 |
| .3 | .330 | .343 | .357 | .370 | .384 | .399 | .414 | .429 | .444 | .460 |
| .4 | .476 | .493 | .510 | .528 | .546 | .564 | .583 | .603 | .624 | .645 |
| .5 | .667 | .689 | .713 | .737 | .762 | .789 | .816 | .844 | .874 | .905 |
| .6 | .938 | .971 | 1.007 | 1.045 | 1.084 | 1.126 | 1.169 | 1.216 | 1.265 | 1.317 |
| .7 | 1.373 | 1.432 | 1.495 | 1.563 | 1.636 | 1.714 | 1.799 | 1.891 | 1.992 | 2.102 |
| .8 | 2.222 | 2.355 | 2.503 | 2.668 | 2.853 | 3.063 | 3.303 | 3.579 | 3.901 | 4.281 |
| .9 | 4.737 | 5.294 | 5.990 | 6.884 | 8.076 | 9.774 | 12.245 | 16.413 | 24.747 | 49.749 |
| 1.0 | $\infty$ | | | | | | | | | |

In practical application, it is often advantageous for the weights to be expressed in small whole numbers, preferably not greater than 10. This permits the multiplying of item-values by weights to be done mentally and then punched into an adding machine. With some loss of precision, numbers such as this may be obtained if each of the weights, as found in the table, is multiplied by the reciprocal of the weight having the greatest magnitude, and then each is rounded to the most similar whole number.

The following examples indicate the necessary operations:

| Factor Loading | Weight | Reciprocal | Transformed Weights | Rounded Weights |
|---|---|---|---|---|
| .78 | 1.992 | .502 | 1.000 | 10 |
| .75 | 1.714 | | .860 | 9 |
| .70 | 1.373 | | .689 | 7 |
| .64 | 1.084 | | .544 | 5 |
| .47 | .603 | | .303 | 3 |

## REFERENCES

1. SPEARMAN, C. *The Abilities of Man.* New York: Macmillan Co., 1927.
2. STEPHENSON, W. *The Study of Behavior.* Chicago: The University of Chicago Press, 1953.

# HISTORICAL INACCURACIES

J. E. WALLACE WALLIN

*311 Highland Avenue, Lyndalia*
*Wilmington, Delaware*

On various occasions certain inaccuracies have been pointed out that have crept into a number of historical reviews that have appeared in psychological and educational journals[11, 12, 13, 14, 15, 17]. "Errors whether unintentional, fictional, or wishfulfilling, or minor or major, if left uncorrected soon become a part of the accepted lore of the profession. Unchallenged errors possess a perseverative tendency and sometimes a tendency to accumulate fictional accretions, just as the rolling snowball gathers anything in its wake that will cling to its surface. Even minor mistakes ought to be expunged from the record for the sake of historical accuracy"[14, p. 21]. The inaccuracies to which reference is here made have included the omission of vital or relevant facts and/or the substitution of fictions for facts. Because of recurrence of vital omissions or misstatements, I am venturing to point out certain errors contained in recitals of the development of certain movements with which I was intimately identified at their inception. Corrigenda relating to clinical psychology and clinical tests appear here; those relating to certain special education developments appear elsewhere[11, 12, 14, 15, 17].

1. It is not a fact that "twenty years ago (*i.e*, in 1917) a group of forty-five men and women holding the doctorate in psychology and engaging in the practice of psychology in the United States held an organization meeting in Pittsburgh to establish the AMERICAN ASSOCIATION OF CLINICAL PSYCHOLOGISTS. Professor Leta S. Hollingworth was largely responsible for initiating the undertaking. Among the group were many well known figures, including Augusta F. Bronner, Arnold Gesell, Henry H. Goddard, William Healy, Rudolph (Rudolf) Pintner, Lewis M. Terman, Lightner Witmer, F. Lyman Wells, and Robert M. Yerkes"[8, p. 23 f.]. I can assure the reader that not more than two of these persons were present at the organization meeting. These accretions remind me of the obscurities and misstatements surrounding the organization of the AMERICAN PSYCHOLOGICAL ASSOCIATION. Even after an exhaustive investigation, involving correspondence with the survivors, Fernberger concedes he is not certain who attended the organization meeting or the initial meeting in Worcester, Massachusetts, on July 8, 1892 [1, p. 35; for a more compbete history; see 2]. Nevertheless, by a process of exclusion her eaches the conclusion that G. Stanley Hall took the initiative in the founding of the American Psychological Association.

Watson's list of 10 or 11 "prominent charter members" published as recently as September, 1953[19, p. 338] does not contain more than two of the bona fide founders of the said Association. Just why the names of the founders should be withheld in these historical recitals is not clear. Who were they? Let us turn to Miner for an historical statement published in 1932, which seems to have been completely overlooked by the historical commentators:

"On December 28, 1917, eight (seven) members of the A. P. A. met at the Carnegie Institute of Technology during the Pittsburgh meeting and organized the AMERICAN ASSOCIATION OF CLINICAL PSYCHOLOGISTS. They elected 47 members of the A. P. A. who accepted membership in the new group. These founders were Leta S. Hollingworth, F. N. Maxfield, J. B. Miner, David Mitchell, Rudolf Pintner, Clara Schmitt, and J. E. W. Wallin. At the third meeting of this organization in Cambridge, Massachusetts, December 31, 1919 . . . the clinical section of the A. P. A. was started and the Association dissolved. Dr. Arnold Gesell acted as chairman from the clinical section and Bird T. Baldwin as chairman from the committee of the A. P. A. . . . Dr. J. E. W. Wallin acted as chairman of the Association of Clinical Psychologists[6, p. 9]." (Parenthetically two of the founders became the first chairmen of the

clinical section—now division—namely, Francis N. Maxfield and David Mitchell). On April 3, 1940, Miner wrote me that he obtained his information from the minutes of the meeting held on December 28, 1917, apparently supplied by Leta Hollingworth who acted as secretary of the meeting.

There was substantial agreement among the founders that some sort of organization should be established to advance the standards of practice in clinical psychology, but violent opposition to the establishment of a new Association developed instanter. In a personal communication under date of December 5, 1953, Dr. Mitchell wrote me that some person whose identity was unknown to him erected a large blackboard notice in the main building of a meeting that amounted to a rump meeting. This notice attracted a large crowd. "In a room which seated about 60 there were nearly as many standing as were seated. The discussion was little less than acrimonious, most carried on by some who disliked the notion of a new organization." I did not see the notice and did not attend that meeting, possibly because I had left the city before the end of the meeting.

It should be mentioned that the organization effected in Pittsburgh was merely provisional or tentative, pending the development of a satisfactory plan of organization, which was to be drawn up by a special committee. Following the Pittsburgh meeting as continuing chairman, I appointed as members of the organization committee Maxfield, Mitchell, and Hollingworth. This committee reported at the second meeting of the organization in Baltimore on December 28, 1918, held in conjunction with the annual A. P. A. convention. So much of the time of this meeting, which attracted a large audience of academic and clinical psychologists, was devoted to the presentation and discussion of the committee report, and the presentation of arguments for and against the organization of a separate association, that little or no time remained for the symposium on clinical psychology especially prepared for the meeting. The symposium, however, was published the following March in the *Journal of Applied Psychology* [3].

It is of interest to state that three of the founders of the precursor of the present Division of Clinical and Abnormal Psychology are still alive, namely, Clara Schmitt, David Mitchell, and the writer. The last two continue as Life Members of the A. P. A.

2. The agitation for the organization of a standardizing, professional association did not have its "inception in 1917 among psychologists in the United States Army during World War I," nor was it "first expressed by members of the American Psychological Association at the time of the World War" [7]. In point of fact, a few clinical psychologists had discussed the advisability of establishing such an organization some five years before our entry into the World War. I distinctly recall discussing the matter in personal contacts with Maxfield, Whipple, Pintner, and possibly others a year or two after I left the New Jersey State Village for Epileptics in 1911, where circumstances made me acutely conscious of the need of professional recognition or status for practicing psychologists [15, 18]. The demand for qualified psychological examiners was precipitated primarily by the criticisms that had been leveled against the inadequate preparation of many Binet testers and "amateur psychologists" who were doing most of the psychological testing and diagnosing of mental deficiency in the public schools of the country at that time [9, limited to the earliest criticisms]. Sixty-three per cent of 115 psychological examiners in the public schools in 1913 were special-class teachers or supervisors or grade teachers, most with only a two-year normal school course. [The Mental Health of the School Child 9, p. 392ff] Many of these Binet testers (the term is not used in a derisive sense) were con stantly referred to in public meetings and in the press as psychologists. The word psychologist had suddenly acquired a new meaning, an extended connotation. Residuals of that connotation still linger, in spite of the exacting current training requirements for clinical psychologists, psychological examiners, and psychoeducational examiners. Similar demands about 40 years ago sometimes evoked ironical, sarcastic rebuffs or bitter denunciations. Some of those episodes that are available in

print—amusing from the present vantage point—will become a part of the complete, unbiased, veridical story of clinical psychology in America if and when it is written. A part of that story appears in The Odyssey of a Psychologist.[15]

One of the incidents that led to the Pittsburgh meeting, I am reminded by Dr. Mitchell, was an informal get-together of a few psychologists at the New York meeting of the A. P. A. in 1916. It was suggested at that time that a few psychologists engaged in clinical practice be invited to attend a roundtable to be held the following year at the time and place of the A. P. A. convention to consider the formation of a professional organization. Mitchell comments that "Dr. Leta Hollingworth was interested in the problem but no more so than several others." She did not take her Ph.D. until 1916, several years after the agitation had begun.

3. Henry H. Goddard was the first to introduce the Binet-Simon scale in America. His translation of the 1905 scale was published in 1908[4] and the 1908 scale in 1910[5]. Both were in use at Vineland before those dates (the actual translator of the scales was, I believe, Elizabeth S. Kite, who died in January, 1954). Guy M. Whipple's translations, which show evidence of the influence of the Vineland translations, did not appear in print until 1910[20]. William Healy, so far as I recall, never published any translation or version of the Binet. If the scale was introduced to this country by Healy "simultaneously" with Goddard[19, p. 323], that fact never came to my attention although I scoured the literature at the time.

4. The true history of the occurrences leading to the introduction of the resolution passed at the Chicago meeting of the A. P. A. in 1915 discouraging the use of mental tests for diagnostic purposes by unqualified examiners is as follows. As part of the campaign I had been waging for several years in addresses and memoirs for the elevation of the competency of Binet testers and psychological testers I prepared a resolution intended for introduction at the business meetings of the A. P. A. and the NATIONAL EDUCATION ASSOCIATION. I went over this resolution at the Chicago meeting of the A. P. A. with Whipple, who was a member of the Council at that time. He agreed to introduce it in as omewhat revamped form[10, 18, and 9, Problems of Subnormality p. 194]. Somewhat later I sent the resolution to a member of the resolutions committee of the NATIONAL EDUCATION ASSOCIATION, which eventually adopted the resolution in the condensed form in which it is reproduced in Clinical and Abnormal Psychology[10].

5. The advocacy of internship training or practicum training did not germinate in the forties. It goes back to 1912 or possibly earlier[16]. Among the greatest advances made in the program of training psychoclinical examiners is the development of supervised internships. But the primary locus for the internship training of school psychologists or psychoeducational examiners should be in selected school systems rather than in hospitals for psychotics. The problems are strikingly different in many particulars.

6. The history of the establishment of public-school special classes for the severely mentally limited—a related problem—has also accumulated accretions of fiction. That phase of historical inaccuracy is given consideration elsewhere[11, 12].

## ADDENDUM

Since this article was written the title of the Division of Clinical and Abnormal Psychology has been changed by action of the A. P. A. to Division of Clinical Psycology. To reference 12 should be added the following article by the same author: The problem of mental retardation. School Executive, 1954, 74, 50-61.

## REFERENCES

1. FERNBERGER, SAMUEL W. American Psychological Association, 1892-1942. *Psychol. Rev.*, 1950, 50, 33-60.
2. FERNBERGER, SAMUEL W. The American Psychological Association: a historical summary, 1892-1930. *Psychol. Bull.*, 1932, 29, 1-89.

3. Gesell, Arnold, Goddard, Henry H., and Wallin, J. E. Wallace. The field of clinical psychology as an applied science. *J. Appl. Psychol.*, 1919, 3, 81-95.
4. Goddard, Henry H. The grading of backward children, the De Sanctis tests and the Binet-Simon tests of intellectual capacity. *The Training School*, 1908, 5, 3-9; see also reprint, August, 1909.
5. Goddard, Henry H. A measuring scale for intelligence. *The Training School*, 1910, 6, 146-154.
6. Miner, James B. The significance and history of the clinical section of the A. P. A. *Psychol. Exchange*, 1932, 1, 7-11.
7. Anon. The proposed American Association of Applied and Professional Psychology. *J. Appl. Psychol.*, 1937, 21, 320-321.
8. Anon. Toward unity. *J. Consult. Psychol.*, 1937, 1, 23-24 (according to the Table of Contents, page III, this anonymous article was written by Percival M. Symonds).
9. Wallin, J. E. Wallace. Clinical psychology: what it is and what it is not. *Science*, 1913, 37, 6-8. Current misconceptions in regard to the functions of Binet testing and of the amateur psychological testers. *Transactions of the Fourth International Congress of School Hygiene*, 1913, 5, 678-688. Danger Signals in clinical and applied psychology. *J. Educ. Psychol.*, 1912, 3, 224-225. *Experimental studies of mental defectives*. Baltimore: Warwick & York, 1912, 109-110. The functions of the psychological clinic. *Medical Record*. 1913, 5-6 (reprint). *The Mental Health of the School Child*. New Haven: Yale University Press, 1914. The new clinical psychology and the psycho-clinicist. *J. Educ. Psychol.* 1911, 2, 207-208. *Problems of Subnormality*. Yonkers: World Book Company, 1917. 106 f., 191-195.
10. Wallin, J. E. Wallace. *Clinical and abnormal psychology*. Boston: Houghton Mifflin Company, 1927, 172.
11. Wallin, J. E. Wallace. Education of children of low ability: corrigenda. *School and Society*, 1954, 80, 119-120.
12. Wallin, J. E. Wallace. The education of severe mental retardates in Delaware and elsewhere. The Sussex Countian, November 5, 1953; reprinted in Education, training and care of mental deficients, a brochure issued by the Sussex Countian, 1953.
13. Wallin, J. E. Wallace. The establishment of the clinical section of the American Psychological Association. *School and Society*, 1938, 48, 114-115.
14. Wallin, J. E. Wallace. Historical authenticity. *J. Except. Children*, 1944, 11, 21-23.
15. Wallin, J. E. Wallace. The *Odyssey of a psychologist*. Lyndalia, Delaware: The Author, 1955.
16. Wallin, J. E. Wallace. Reaverments regarding psychoclinical norms and scales of development. *Psychological Clinic*, 1913, 3, 89.
17. Wallin, J. E. Wallace. The school psychologist in retrospect. *J. Consult. Psychol.*, 1942, 6, 309-312.
18. Wallin, J. E. Wallace. Vagrant reminiscences of an oligophrenist. *Amer. J. Mental Deficiency*, 1953, 58, 39-55.
19. Watson, Robert I. A brief history of clinical psychology. *Psychol. Bull.*, 1953, 50, 321-346. (Comprehensive as it is this review needs to be supplemented by a comprehensive history of the development of texts in clinical psychology.)
20. Whipple, Guy M. *Manual of mental and physical tests*. Baltimore: Warwick & York, 1910, 473-517.

# EDITORIAL OPINION

## A NEW APPROACH TO THE PROBLEM OF THE HIERARCHICAL ORGANIZATION OF PERSONALITY INTEGRATION

Clinical psychologists everywhere should pause to give consideration to an accumulating body of research evidence questioning the validity of current standard practices as applied to the making of clinical predictions and decisions. It does not seem unfair to state that our current methods of analysing and interpreting data from life history, self-ratings, behavior ratings by others, objective tests and projective techniques are of questionable validity in relation to the types of predictions and decisions which we attempt to make from them. The pioneer investigations of Kelly and Fiske[1] on the prediction of performance in clinical psychology should have been interpreted as a danger signal warning us of the lack of validity of current methods for making socially significant predictions about behavior. Subsequent studies, such as those of Holsopple and Phalen[2] and Holtzman and Sells[3], indicate that using the best of available methods, the predictions of many clinicians are not significantly different from chance expectations and that frequently when the clinician is most certain of the validity of his judgments he is most apt to be in error. Indeed, the available data indicate great variations in the validity of predictions by different clinicians, with some clinicians performing considerably worse than chance. The general conclusion to be drawn from this trend of evidence is that many current methods and diagnostic procedures are not valid for the purposes for which they are being used. With greater availability of *validating uses* of statistical predictions, it is no longer tenable for the clinician to avoid empiric validation of his clinical judgments and predictions.

There are many theoretical reasons for the practical failure of current testing practices to prove valid. First, in many instances testing practices have been founded on speculative theories which are themselves unvalidated, *e.g.* many of the practical applications of psychoanalytic theory. Second, there has been a general failure to identify and objectify *all* of the variables underlying the phenomena under study. In many instances, complex phenomena (such as schizophrenia or psychoneurosis) which are actually unanalyzed, have been treated as simple variables whose composition was actually known. Third, there has been a general failure to discriminate between etiologic factors which underly socially significant variables and those which have no important predictive value. Many tests have been reliable and valid indices of factors which have no social significance. Fourth, there have been great technical difficulties in establishing satisfactory criteria against which to validate test items. Lacking valid formulations of etiologic relationships, it has been necessary to utilize criteria of hypothetical validity which have not always been authenticated. Fifth, there has been a tendency to utilize single tests for multidimensional measurements far beyond any valid rationale concerning what could be validly expected from them particularly in the field of projective testing. Sixth, it has only been recently that factorial validation has become possible through identification of factors and marker variables. This development promises to revolutionize existing methods. Finally, existing tests make possible only a very incomplete sampling of the wide variety of factors which actually organize higher behaviors involved in the organism coping with complex environmental situations.

---

[1]Kelly, E. L. and Fiske, D. W. *The Prediction of Performance in Clinical Psychology.* Ann Arbor, Mich.: University of Michigan Press, 1951.

[2]Holsopple, J. Q. and Phelan, J. G. The skills of clinicians in analysis of projective tests. *J. clin. Psychol.*, 1954, 10, 307-320.

[3]Holtzman, W. H. and Sells, S. B. Prediction of Flying Success by analysis of test protocols. *J. abn. soc. Psychol.*, 1954, 49, 485-490.

Your editor is not personally surprised over this trend of events since for more than 20 years we have been very critical of practices which have involved the invalid use of methods far beyond anything which is scientifically known about them. We have been dismayed by the lack of any systematic foundations for clinical practice based solidly on basic psychological science. While many clinicians have become very proficient in the use of single techniques, few have acquired any genuinely eclectic orientation to *all* levels of organization of personality integration. The total complexity of the problem of psychological examining has not been widely faced with all its implications for a comprehensive system of psychodiagnosis. Psychological examining will continue to be invalid in relation to clinical predictions and decisions until it actually becomes able to deal with all the factors which organize socially important behaviors.

It was within this frame of reference that your editor began about 1940 to formulate the basic principles of a comprehensive system of psychological examining which would encompass all levels of behavior organization known to underly socially significant personality traits. Approaching this endeavor from a psychobiological viewpoint emphasizing the evolution of behavior both in the species and in the person, it soon became evident that the central problem was one of the hierarchical organization of personality integration starting with the lowest and most primitive psychophysiological foundations and working up to the highest level behaviors of which man is capable. We began by formulating an hierarchical level theory of organization of personality integration, seeking to identify the etiologic factors operating at all levels of personality integration in order to achieve an operational understanding of the methods for diagnosing the actual determiners of socially significant behaviors at all levels of integration. Instead of starting with the common approach of listing current tests and methods and seeing what can be gotten from them, we started with a systematic theory of the organization of behavior at all levels of integration and then seeking to discover how they could be studied with existing methods. This was accomplished by arranging an outline or classification of statistically derived *factors* and then listing the operational methods available for their study. The central problem in all of psychological examining is recognized to be the identification of the etiologic factors *actually organizing* socially significant behaviors. The purpose of this editorial is to announce the completion and publication of this system of psychological examining in a textbook soon to appear.[4] Our effort has been deliberately patterned after the work of Sir William Osler who pioneered at Johns Hopkins University at the turn of the century to establish a uniform system of medical practice to replace the conglomeration of medical cults and proprietor-owned medical schools which characterized prescientific medical education and practice. Whether we have succeeded or not will be left for history to decide, but at least an attempt has been made to establish the first comprehensive and eclectic system of psychological examining and therapy.

[4] THORNE, F. C. *Principles of Psychological Examining.* Brandon, Vt.: Journal of Clinical Psychology, 1955, in press.

F. C. T.

# BOOK NOTICES

DEUTSCH, F. and MURPHY, W. F. *The Clinical Interview. Vol. I: Diagnosis.* New York: International Universities Press, pp. 613. $10.00.

The authors are practising psychiatrists with long experience in psychoanalytically-oriented training programs for psychiatric residents. A subtitle to the book describes their purpose of outlining a method of teaching associative exploration which involves the attempt to trace specific interrelations in depth and to elucidate the structure of symptom formation. Following two introductory chapters, the main part of the book consists of verbatim recordings of psychoanalytic type interpretive interviews with comments to elucidate the dynamics of interpretive interviewing.

BECK, S. J. The Six Schizophrenias. *Amer. Orthopsychiat. Assoc. Monogr.*, 1954, No. 6, pp. 238. $5.00.

This is a report of a five year research project on the subject of schizophrenia subsidized by the National Mental Health Institute and carried out by a team at the Michael Reese Hospital, Chicago. Integrating the results from Rorschach analyses and Q-technique analyses of trait universes derived from psychiatric definitions of schizophrenia, the authors were able to differentiate six patterns of schizophrenia which are defined operationally in terms of variables having high loading for each factor. The methodology of this research marks a great advance in techniques for analysing symptom complexes.

LOWENFELD, M. *The Lowenfeld Mosaic Test.* London: Newman Neame, Ltd., 1954, pp. 360 with 144 colored plates in a separate container. Sold in the United States by the Psychological Corporation, 522 Fifth Ave., New York 36, N. Y. $10.00. The Lowenfeld Mosaic materials are also available at $27.00 for a single set, or $43.50 for a double set.

The author is founder and physician-in-charge of the Institute of Child Psychology, London. This book reports the results of 20 years of research with the Lowenfeld Mosaic materials which consist of 456 pieces of colored plastic in five interrelated geometric shapes each presented in six colors. Clinical experience indicates that personality characteristics are reflected in designs and selections of colors made. Part I consists of four chapters describing the origin of the test, principles of classification of responses involving abstract patterns with and without recurring form. Part II discusses clinical findings with groups of children, mental defectives, normals, neurotics, psychotics and in the study of cultural problems.

THRALL, R. M., COOMBS, C. H. and DAVIS, R. L. (Eds.) *Decision Processes.* New York: John Wiley, 1954, pp. 332.

This book reports empirical research and theory construction growing out of an eight week seminar on The Design of Experiments in Decision Processes held under the auspices of the Ford Foundation in 1952 at Santa Monica, Calif. Two introductory chapters discuss the relation of mathematical models to measurement theory. Part I includes five chapters on individual and social choice; part II includes three chapters on learning theory with respect to multiple choice situations, uncertain situations and game-learning theory; part III presents the theory and applications of utility; and part IV describes some experimental studies.

MCNEMAR, Q. *Psychological Statistics. 2nd Ed.* New York: John Wiley, 1955, pp. 408. $6.00.

MEEHL, P. E. *Clinical vs. Statistical Predictions* Minneapolis, Minn.: University of Minnesota Press, 1954, pp. 149. $3.00.

This monograph presents a theoretical analysis and review of the evidence concerning the validity of clinical vs. statistical predictions. The conclusion is reached that the validating use of statistics is unavoidable in order to demonstrate whether clinical predictions are significantly different from chance probabilities.

HOLSOPPLE, J. Q. and MIALE, F. R. *Sentence Completion.* Springfield, Ill.: C. C. Thomas, 1954, pp. 177. $5.50.

This monograph reports the development and clinical application of a 73-item Sentence Completion test. Illustrative records are given in detail but there is no formal standardization. The price seems excessively high for a monograph this size.

ALLEN, R. M. *Elements of Rorschach Interpretation.* New York: International Universities Press, 1954, pp. 242. $4.00.

Along with the author's earlier publication, *Introduction to the Rorschach Technique: Manual of Administration and Scoring,* this book is offered as an elementary text suitable for a one-year course in basic Rorschach.

BLUM, L. H., DAVIDSON, H. H. and FIELDSTEEL, N. D. *A Rorschach Workbook.* New York: International Universities Press, 1954, pp. 169. $2.00.

This book is an introduction to Rorschach scoring presenting over 1000 self-testing exercises based on actual responses given by subjects. The student will find this a valuable adjunct to laboratory practice in Rorschach technique.

FERM, V. *A Dictionary of Pastoral Psychology.* New York: Philosophical Library, 1955, pp. 336. $6.00. Definitions of psychological terms relevant to ministers.

EISSLER, R. S. *et al* (Eds.) *The Psychoanalytic Study of the Child. Vol. IX.* New York: International Universities Press, 1954, pp. 369. $7.50.

EISENSON, J. *Examining for Aphasia.* New York: Psychological Corporation, 1954, pp. 73. Revised edition.

HAYES, E. N. *Directory for Exceptional Children.* Boston: Porter Sargent Publishers, 1954, pp. 96. $1.10. A listing of private schools, tutors, summer camps, psychiatric and guidance clinics, state services, state schools, associations, societies and periodicals devoted to problems of exceptional children.

DURKIN, H. E. *Group Therapy for Mothers of Disturbed Children.* Springfield, Ill.: 1954, pp. 125. $3.50.

LEWIS, H. *Deprived Children.* New York: Oxford University Press, 1954, pp. 163. $1.55.

This monograph reports a ten year study supported in England by the Nuffield Foundation on problems relating to the reception, investigation and disposal of deprived children. A study was made of 500 children referred to a typical county reception center of which followup studies were made of 240 to discover their condition two years after their reception. The reception center plan for the care of deprived children was enacted into law in the Children's Act of 1948 requiring every county authority to establish at least one such center. Social welfare authorities everywhere should become familiar with the results of this experiment.

WOLFENSTEIN, M. *Children's Humor.* Glencoe, Ill.: The Free Press, 1954, pp. 224. $3.75.

GENGERELLI, J. A. (Ed.) *The Psychological Variables in Human Cancer.* Berkeley, Calif.: University of California Press, 1954, pp. 135. $3.00. Reports of a symposium held at the Long Beach Veterans Administration Hospital in 1953.

TAYLOR, W. S. *Notehand for Psychologists.* 15 Pierce Hall, Northampton, Mass.: mimeographed, 1954, pp. 5. $.21 in stamps.

This system abbreviates most of the 100 words that are more than 50% of ordinary writing, many words almost as common, and many general categories, technical terms, and common letter-groups. The system is well tested, is easily learned, and reduces ordinary writing about 30%, and psychological writing, 35%.

**THE SOUTHARD SCHOOL**

**Intensive individual psychotherapy in a residential school, for children of elementary school age with emotional and behavior problems.**

**THE MENNINGER CHILDREN'S CLINIC**

**Outpatient psychiatric and neurologic evaluation and consultation for infants and children to eighteen years.**

*Department of Child Psychiatry*

**THE MENNINGER FOUNDATION**

**J. Cotter Hirschberg, M. D., Director** **Topeka, Kansas; Telephone 3-6494**

**Essential facts and theories plus a basis for future research**

THEORIES OF PERCEPTION AND THE CONCEPT OF STRUCTURE

**A Review and Critical Analysis**
**with an Introduction to a Dynamic-Structural Theory of Behavior**

*By* FLOYD H. ALLPORT, *Syracuse University*. This unique, dual-purpose volume provides the first comprehensive, organized survey of the major current theories of perception. It also offers a foundation for future systematic work in the form of a general theory of structure in behavior.

The work opens with a review of the essential facts and theories of perception, critically examining them in the light of scientific method. It then covers each of the current major theories separately, bringing out their strong and weak points and correlating them one with another. From this, the author develops a synthesis of the major generalizations more or less common to all theories.

From this examination the need for a new and general theory of structure in perception and behavior studies emerges clearly. To answer it, the author proposes an over-all, explicit theory of structure. He then re-examines the various theories in the light of his new theory.

**1955. 709 pages. $8.00.** ***Send now for a copy on approval.***

**JOHN WILEY & SONS, Inc., 440 Fourth Avenue, New York 16, N. Y.**

# *d is for discipline*

# *D is for Devereux Schools*

where the multi-disciplined approach is carried through to its maximum potential in dealing with the problems of childhood and adolescence.

The Devereux Plan is a never-ending fusion of disciplines—educational, medical, psychiatric, and psychological—not only in diagnostic work, but in every conceivable aspect of BOTH staff and student experience.

This, in turn, is made possible by a full score of geographically separated, wholly autonomous, school-home units—enabling optimum effective placement in terms of both personality structure and diagnostic categories. Equal but separate facilities are available for children with "learning disabilities"; conflictual youngsters of average or superior intelligence; and other groups, at various age levels. Thus, each group can profit in its own way from a fusion of professional disciplines within a dynamically oriented framework.

We invite you to see for yourself, at any time, what the Devereux Plan is accomplishing. Or, if you wish written information, you are invited to write to

JOHN M. BARCLAY, DIRECTOR OF DEVELOPMENT,
DEVEREUX SCHOOLS, DEVON, PA.

---

THE DEVEREUX FOUNDATION
CALIFORNIA ..... PENNSYLVANIA

## Devereux Schools

UNDER THE DEVEREUX FOUNDATION

HELENA T. DEVEREUX, *Director*
J. CLIFFORD SCOTT, M.D., *Executive Director*

SANTA BARBARA, CALIFORNIA — DEVON, PENNSYLVANIA